Journal of
Experimental Psychology

VOL. 26, NO. 1 JANUARY, 1940

FUNDAMENTAL PROBLEMS IN COLOR VISION. II. HUE, LIGHTNESS, AND SATURATION OF SELECTIVE SAMPLES IN CHROMATIC ILLUMINATION

BY HARRY HELSON AND VIRGINIA BALOUGH JEFFERS

Bryn Mawr College

That the limits of color constancy are reached in everyday experience is shown by the fact that colors of objects do not remain constant in different illuminants and when various conditions of viewing are changed. Thus a suit of clothes which is a good bluish gray in the lamplight of the store may be a saturated blue in daylight. Changes in color which escape the layman are easily detected by experts engaged in matching and grading objects by their colors.

The belief is widely held that changes in object colors in strongly chromatic illuminants, when the whole retina is stimulated by homogeneous light, are either wholly inexplicable or unamenable to the laws governing normal color vision. The conflicting hypotheses advanced for the facts of constancy and conversion have not contributed to a clearer understanding of the phenomena or their underlying mechanisms. The application of laboratory controls to the phenomena of color conversion has resulted in the formulation of principles valid for non-selective samples in various illuminants.[1] As most

[1] D. B. Judd, Surface color, *J.O.S.A.*, 1935, **25**, 44; H. Helson, Tri-dimensional analysis and the non-film modes of color appearance, *J.O.S.A.*, 1937, **27**, 59, and Fundamental problems in color vision. I. The principle governing changes in hue, saturation, and lightness of non-selective samples in chromatic illumination, this *Journal*, 1938, **23**, 439–476.

objects have chromatic daylight colors, that is, possess spectrally selective reflectance, the question arises: "Do the laws for non-selective samples hold for selective ones as well?" The present paper deals, therefore, with the facts and principles of changes in hue, lightness, and saturation of selective samples in spectrally homogeneous and mixed illuminants.

Procedure and Experimental Controls

As a full account of the apparatus and technique of experimentation has already appeared in a previous publication of this *Journal*[2] we shall limit ourselves to the most essential facts. For univocal answers to the questions which have concerned us it was necessary to use illuminants of known spectral energy distributions, stimuli and backgrounds of known spectral reflectance, and subjects (S) trained to report on hue, lightness, and saturation with fair consistency.[3] The following arrangements contributed to these ends:

The source of illumination was a 500-watt Mazda lamp equivalent to a black body radiator of 2848° K. when operated at 103 volts. This source lamp reproduces 'Illuminant *A*' recommended by the International Commission on Illumination.[4] It was used chiefly in conjunction with spectrally selective filters to give strongly chromatic illumination. Some observations, however, were taken without any filter as representative of weakly chromatic illumination and these were supplemented by observation with a Palo Daylight lamp which gives illumination that is nearly achromatic or 'white.' Observations were made in a light-tight booth to guard against stray light which can materially affect results.

The transmissions of the filters for 2848° K., the trilinear coördinates of the illuminants,[5] the illumination on the sample plane, with and without filters interposed, the hues of the illuminants, and their after-image complementaries are given in Table 1. From this table it will be seen that transmissions of the filters, and hence the illumination on the samples, differed by a factor of 60 for each intensity of source which we employed. This range of illuminations was counterbalanced by deliberate change in source intensity by a factor of 96, that is, by varying the amount of light from the source it was possible to produce 96-fold change in illumination with each filter. Actual intensities from the source were 135, 72, 4.5, and 1.4 foot-candles on the sample plane and when multiplied by the transmissions of the filters as given in Table 1 we have

---

[2] H. Helson, *ibid.*

[3] We are grateful to Dr. Deane B. Judd of the National Bureau of Standards for advice on experimental techniques, aid in measurement of spectral reflectances and transmissions, and for his unfailing interest in the progress of the work in all its stages. We trust that our debt to him has been repaid at least in part by some of the results of our experiments which he has built into the theory and formulæ which appear under his own name.

[4] *Cf.* D. B. Judd, The 1931 standard observer and coördinate system for colorimetry, *J.O.S.A.*, 1935, **25**, 24–35.

[5] For definitions of various terms especially from colorimetry see Judd, *ibid.*, and Helson, *ibid.*

the illumination on the samples for each source-filter combination. These products are given in Table 1 for the main case, *viz.*, 72 f.c. on the sample plane.

The hues of the illuminants, as revealed by reports on the colors of the lightest non-selective samples, were not all unitary, that is, they require more than one hue-name to describe them.[6] It should be noted that we have indicated in Table 1 the hues of the after-image complementaries to the illuminant hues *because our results are explicable only in terms of after-image hues and not in terms of mixture complementarism.* The former complementaries differ from the latter in that they contain more red and blue (purple) than the latter.[7]

TABLE 1

TRANSMISSIONS OF THE CHROMATIC FILTERS, ILLUMINATION ON THE SAMPLE PLANE, TRILINEAR COORDINATES OF THE ILLUMINANTS, HUES OF THE ILLUMINANTS, AND THEIR AFTER-IMAGE HUES

	No Filter	Red	Yellow	Green	Blue
Transmission for 2848° K.....	1.000	0.052	0.362	0.038	0.006
Foot-candles on samples.....	72.00	3.70	26.00	2.70	0.44
Trilinear coördinates.....	r g	r g	r g	r g	r g
	.543 .424	.958 .042	.705 .295	.372 .613	.237 .351
Hue of illuminant	Y of low saturation	yR	rY	yG	rB
After-image hue..	Indeterminate low saturation	bG	B	bR	Y

The color stimuli consisted of 15 Munsell selective samples, 50 × 50 mm as given in Tables 2 and 5, and three non-selective samples identical with our black, gray and white backgrounds, making 18 stimuli in all. Spectral reflectance curves of the 15 Munsell samples were determined at the National Bureau of Standards and the trilinear coördinates of each sample in each illuminant were calculated. These data are re-

[6] The yR illuminant is in effect spectrally homogeneous even though it is yellowish-red because all of the samples illuminated by it fall closely on the same part of the spectrum locus when plotted in the color mixture diagram. The reddish-yellow, yellowish-green, and reddish-blue illuminants do not plot on the spectrum locus and are therefore equivalent to non-homogeneous lights. In laying claim to the use of homogeneous illuminants we shall have only the first in mind. It should be remembered that purity of stimulus does not give unitary hues always—in fact the reverse is true as only three psychologically unitary hues have been found in the spectrum—G, Y, and B, all reds in the spectrum having a yellow component. For most recent determination of spectral unitaries, see F. L. Dimmick and M. R. Hubbard, The spectral location of psychologically unique yellow, green and blue, *Am. J. Psychol.*, 1939, **52**, 242–254.

[7] A. Tschermak, Licht und Farbensinn, *Handb. d. norm. u. pathol. Physiol.*, Berlin, 1929, XII/1, Receptionsorgane 11, pp. 474 f.; also T. Karwoski, Variation toward purple in the visual after-image, *Amer. J. Psychol.*, 1929, **41**, 625–636.

served for a later article but the reflectance of each sample for each illuminant is given in Tables 2 and 5. Since the lightness of a sample is directly dependent upon its reflectance this value is of prime importance.

The Ss in these experiments numbered 21. Sixteen observed during the first year of experimentation and 5 observed during the major variations carried on during the second year, two of the five having observed in both years' work. Ss were trained to report on hue, lightness, and saturation in daylight by the use of the Munsell color book, after which their reports in the color booth were accepted at their face value. Reports on hue were made in terms of a sixteen-fold classification, R, yR, RY, rY, Y, gY, YG, yG, G, bG, BG, gB, B, rB, RB, bR, the small letters denoting minor components in binary hues. The abbreviations are self-evident for the hues and the achromatic series was denoted by A 10/0, A 5/0 and A 0/0 for white, gray, and black. The Munsell notation was used wherein the number in the numerator refers to lightness and the one in the denominator to saturation. Maximum saturation or lightness was denoted by 10 and minimum by 0. Thus a black would be denoted by A 0/0 and a yellowish-red of medium saturation and lightness by yR 5.0/5.0.

Each S adapted to the illuminant for 15 minutes before the samples were exposed. S was instructed to glance about the field without fixating too long on any one sample as steady fixation often caused the samples to change in color. Under our conditions the whole retina was stimulated with chromatic illumination. The total range of conditions investigated in this study and the number of observations on which our conclusions are based can be gained from the following: 18 samples were reported on by 5 Ss (whose results we shall concentrate on in the present report) in 25 different illuminants (incandescent, yR, rY, yG, rB and mixtures of the chromatic illuminants with the lamp light), with three backgrounds (white, gray and black) in four intensities of each illuminant (high, medium, low, and very low), resulting in a total of 27,000 ($18 \times 5 \times 25 \times 3 \times 4$) reports on hue, lightness, and saturation.

General Characteristics of Colors in Strongly Chromatic Illuminants

When we view a number of differently colored papers in illumination having a continuous spectrum, we see a variety of colors in every sense of the word. The reds stand out from the greens while the blues and yellows appear in their individuality also. Various degrees of lightness and saturation add to the variety of hues in the field. Multiplicity of colors characterizes the ordinary conditions of vision. The colors are solid as the glance falls upon them and they stand out from their background. They are one with the objects to which they belong. Some colors are more pleasing than others and the affective value of any color can be isolated more or less from the rest. Or if we regard the colors as a whole both the color-qualities as such and their affective values seem to adhere to the objects.

If, however, we change the illuminant from daylight to any strongly chromatic illuminant, the experience changes in very radical fashion. Objects and their colors lose their individuality to a great extent. A single hue and mood pervade the field. There is a peculiar luminous quality to everything that is seen. Colors retreat into the plane of the background, some even appearing within or behind it. The observer seems bathed in the illuminant hue. Wherever one looks he is conscious of the illuminant hue even though differences in color are visible. The colors have become softer and filmier. Demarcation between samples and background is not sharp and contours almost disappear. Variety, individuality, and thingness give way to a single predominating color tone wherein objects seem fused. The affective value of the daylight plane has changed to include three-dimensional space. Affective tone is stronger than in ordinary light and one is more conscious of the effect of color on feeling and mood than in normal surroundings.

One of the most striking differences between daylight hues and those seen in strongly chromatic illuminants is the loss in saturation which the latter have undergone. It is most pronounced in those samples having a daylight hue like that of the illuminant and seems paradoxical at first. In spite of the fact that 'red' samples, to take this hue as an example, reflect a greater fraction of homogeneous red light than daylight, and similarly for samples of other hues, they have a very much less saturated color in strongly chromatic red illumination owing to the desensitization of the eye to light of the illuminant hue.[8] Two factors are at work: strong stimulation by certain spectral energies to produce a given hue and retinal adaptation to neutralize this effect. Thus it is possible for samples having

[8] For this reason when maximum values are given for the saturation of illuminant hues it does not mean that the samples are as saturated as possible or even as saturated as in daylight. Both A. Gelb, Die Farbenkonstanz der Sehdinge, *Handb. d. norm. u. pathol. Physiol.*, Berlin, 1929, XII/1, Receptionsorgane 11, pp. 624 ff., and D. Katz, Die Erscheinungsweisen der Farben und ihre Beeinflussung durch die individuelle Erfahrung, *Z. F. Psychol.*, 1911, Erg. Bd. 7, pp. 270 ff., have found that samples having the illuminant hue are less saturated than in daylight illumination. But the complementary hues are fully as saturated at their best as any seen in daylight.

a daylight hue different from that of the illuminant to be more saturated than samples like the illuminant—for example, a daylight white (having higher reflectance) is redder than daylight reds in red illumination. On the other hand, the after-image hues are more saturated, stronger, and deeper than the illuminant hues and appear equally well in any samples regardless of daylight hue provided reflectance is sufficiently low for the illuminant.

The after-effects of exposure to strongly chromatic illumination are also striking. Immediately after the illuminant is changed from strongly chromatic to daylight, colors are greatly desaturated for many seconds, especially those having the hue of the chromatic illuminant. This 'after-desaturation' was greatest with the yR and rB illuminants and least with the yG and rY. This finding is in harmony with the fact that the red and blue-violet portions of the spectrum show greater fatigability than the yellow and green regions even though their saturation is greater.[9] Since the yR and rB illuminants plot farthest from the daylight point in the color mixture diagram we expect them to have the greatest chromatic effects and after-effects.

Hue, Lightness, and Saturation in Chromatic Illuminants

In attempting to formulate principles governing color conversion it is necessary to use an average, median or some other typical S as the object of predictions owing to individual differences in color vision. Unanimity of report on colors is not found even under optimal daylight conditions and so it cannot be expected when observations are much more difficult as in chromatic illumination. The chief types of variation in

[9] According to D. McL. Purdy, On the saturations and chromatic thresholds of the spectral colors, *Brit. J. Psychol.*, 1930–31, **21**, 303, violet requires minimal intensity for maximal saturation, red is next, and greenish-yellow (565 mμ) needs maximal intensity for full saturation. He points out that the highly saturated colors require the minimal *time* to desaturate and lose their saturation first as intensity is raised—findings which we have corroborated under entirely different conditions in our after-desaturation effect and in the return to daylight hue with admixture of lamp light—red and blue requiring less than yellow, the exception of green being discussed below.

report found in chromatic illuminants which have also been reported among Ss with so-called normal color vision in daylight [10] are as follows:

(1) Samples which have unitary hues for the majority of Ss have binary hues for the remaining Ss—for example, Y 5/7 reported Y by 10 Ss was seen as gY by 6 others.

(2) Components of binary hues may be given different emphases: thus the sample P 3/6 was called RB by 5 Ss, rB by 6, and bR by 5.

(3) Minor components of binary hues may be on either side of the major component. The sample 7/6 which was called R by 13 Ss was seen as yR by 2 and bR by one S.

(4) Less agreement exists regarding unitary hues (hues supposed to be unitary) than with respect to binary hues as would be expected from the above. But binary hues are seldom seen as unitary since the tendency is to add components rather than to miss them.

By using illuminants of known composition and papers of known spectral reflectance and by excluding such sources of error as arise from stray light, insufficient adaptation, and inexperience in naming colors, the general principles governing conversion emerge concretely and clearly through the individual differences in color vision.

Consideration of the data in Table 2 for the median *S* reveals certain regularities. If we arrange the samples in order of decreasing reflectance for each illuminant we find first reports of illuminant hue or mixtures of daylight and illuminant hues in the case of the less homogeneous illuminants (rY and yG), secondly, reports of achromaticity, which may not be present in all cases, and thirdly, on white and gray grounds, after-image hues with the darkest samples. Taking the yR illuminant as an example, on white ground the first 10 samples are R or YR, the next sample is A, and the remaining 7 samples are, with one exception, GB; on gray ground, the first 12 samples are RY, the next two are YG and Y, and the remaining four are GB or G; on black ground, all the

[10] H. Helson, Color tolerances as affected by changes in composition and intensity of illumination and reflectance of background, *Amer. J. Psychol.*, 1939, **52**, 406–410.

TABLE 2

HUE (H), LIGHTNESS (L), AND SATURATION (S) FOR THE MEDIAN S, AND APPARENT REFLECTANCE (A) OF 18 MUNSELL SAMPLES (M), IN yR, rY, yG, AND rB, ILLUMINANTS ON WHITE, GRAY AND BLACK BACKGROUNDS

Yellowish-Red Illumination						Reddish-Yellow Illumination					
M	*A*	White	Gray	Black		*M*	*A*	White	Gray	Black	
		H *L/S*	*H* *L/S*	*H* *L/S*				*H* *L/S*	*H* *L/S*	*H* *L/S*	
A 10/0	.800	R 9.0/4.0	RY 8.0/7.5	yR 8.0/9.0		A 10/0	.800	Y 9.0/4.5	Y 8.5/9.0	Y 9.0/8.0	
R 7/6	.781	yR 9.0/3.0	RY 8.0/7.0	yR 7.5/9.0		R 7/6	.618	RY 8.0/5.0	rY 7.5/7.0	rY 8.0/7.5	
R 5/10	.706	yR 8.0/3.5	yR 8.0/7.0	yR 7.5/9.0		Y 7/8	.588	gY 8.0/4.0	Y 8.0/7.0	Y 8.0/7.0	
Y 7/8	.577	RY 8.0/3.0	yR 7.5/6.0	yR 6.5/9.0		B 7/4	.445	yG 7.0/5.0	YG 8.0/5.5	YG 7.0/5.5	
pB 7/4	.428	yR 7.0/2.5	RY 7.0/4.5	R 6.0/8.0		G 7/7	.428	yG 7.0/7.0	yG 7.5/7.0	yG 7.0/5.0	
B 7/4	.381	rY 6.0/1.0	rY 6.0/5.0	R 5.5/8.0		pB 7/4	.405	gY 7.0/3.0	gY 5.0/4.5	gY 7.0/5.0	
P 5/6	.359	yR 6.0/2.0	RY 6.0/5.0	R 5.5/8.0		R 5/10	.404	yR 7.0/8.0	yR 7.0/9.0	RY 7.0/8.5	
G 7/7	.329	R 5.0/1.0	rY 6.0/4.0	yR 5.0/7.0		Y 5/7	.288	rY 5.0/1.0	gY 5.0/3.5	yG 6.0/5.0	
YR 4/5	.297	yR 4.5/1.0	RY 5.0/4.0	yR 4.5/7.0		P 5/6	.273	bR 5.0/3.5	RY 5.0/3.5	rY 6.0/4.0	
R 3/7	.276	R 4.5/1.0	RY 5.0/4.0	R 4.5/7.0		YR 4/5	.246	bR 4.5/3.5	rY 5.0/3.5	rY 5.5/4.0	
Y 5/7	.272	A 4.5/0.0	rY 5.0/4.0	R 4.5/7.0		A 5/0	.240	gY 4.5/0.5	Y 5.0/1.5	Y 6.0/2.0	
A 5/0	.240	R 4.0/0.4	yR 5.0/1.5	R 4.5/3.5		B 5/6	.185	bG 3.0/0.5	yG 3.0/3.0	yG 5.0/4.0	
B 5/6	.161	gB 3.0/1.0	YG 3.0/0.5	R 3.0/5.0		G 5/7	.171	bG 4.0/8.0	G 5.0/9.0	G 5.0/8.5	
P 3/6	.153	gB 3.5/1.0	Y 3.0/0.7	bR 3.0/5.0		R 3/7	.151	bR 4.5/9.0	bR 5.0/8.5	yR 5.0/7.0	
G 5/7	.087	GB 1.0/4.0	gB 2.0/2.0	yR 2.0/2.0		P 3/6	.119	bR 3.0/5.0	yR 3.0/4.0	RY 4.5/4.0	
G 3/4	.072	GB 1.0/7.0	GB 2.0/2.0	yR 1.5/2.0		G 3/4	.078	rB 2.0/2.5	rB 2.0/2.0	G 3.0/2.0	
B 3/5	.064	bG 2.0/7.0	BG 1.5/2.0	yR 1.5/2.0		B 3/5	.072	rB 2.0/2.5	rB 2.0/2.0	yG 3.0/3.0	
A 0/0	.030	gB 2.5/9.0	G 2.5/9.0	A 0.0/0.0		A 0/0	.030	rB 1.0/2.0	RB 1.0/3.0	A 0.0/0.0	

TABLE 2—*Continued*

Yellowish-Green Illumination									Reddish-Blue Illumination							
M	*A*	White		Gray		Black			*M*	*A*	White		Gray		Black	
		H	*L/S*	*H*	*L/S*	*H*	*L/S*				*H*	*L/S*	*H*	*L/S*	*H*	*L/S*
A 10/0	.800	G	8.5/3.0	gY	9.0/9.0	gY	8.5/7.0		A 10/0	.800	B	9.0/3.5	B	9.0/6.0	B	8.5/7.5
G 7/7	.596	yG	8.0/3.0	YG	8.0/5.0	yG	7.5/6.0		B 7/4	.642	B	8.0/6.0	B	8.5/6.0	B	8.0/6.0
B 7/4	.581	yG	8.0/3.0	yG	8.0/5.0	yG	7.5/6.0		pB 7/4	.600	B	8.0/5.0	B	8.5/7.0	B	8.0/6.0
pB 7/4	.485	G	7.0/3.0	yG	6.5/6.0	yG	7.0/6.0		G 7/7	.493	gB	7.0/4.5	B	7.0/6.0	B	7.5/5.0
Y 7/8	.455	rY	6.5/4.5	Y	6.5/5.0	gY	7.0/4.5		B 5/6	.472	gB	7.0/3.5	B	7.5/7.0	B	7.5/8.0
G 5/7	.377	yG	5.0/2.5	yG	6.0/7.0	yG	6.0/7.0		R 7/6	.346	rB	6.0/6.0	RB	6.0/5.0	rB	6.5/6.0
R 7/6	.357	Y	5.5/2.5	Y	6.0/3.0	YG	6.0/4.0		P 5/6	.310	rB	6.0/6.0	rB	6.5/7.0	rB	6.0/8.0
B 5/6	.324	bG	5.0/3.0	bG	5.0/7.5	G	5.0/7.0		G 5/7	.267	G	6.0/7.0	G	6.0/7.0	BG	5.5/6.5
A 5/0	.240	A	5.0/0.0	gY	5.0/2.0	yG	5.0/2.0		A 5/0	.240	A	5.0/0.0	A	5.5/0.0	B	5.0/5.0
P 5/6	.235	rB	5.0/2.0	YG	5.0/0.5	yG	5.0/5.0		B 3/5	.217	B	5.0/0.5	BG	5.0/1.0	B	5.0/6.5
Y 5/7	.221	RY	5.0/4.0	rY	5.0/4.0	yG	5.0/3.0		Y 7/8	.159	rY	3.0/3.5	rY	4.0/3.0	RB	4.0/1.5
B 3/5	.133	rB	3.0/4.0	B	3.0/1.0	G	4.0/7.0		P 3/6	.153	RB	5.0/6.5	rB	4.0/7.5	RB	4.0/8.0
R 5/10	.110	bR	3.0/5.0	bR	4.0/4.0	gY	4.0/2.0		R 5/10	.110	bR	3.0/1.5	bR	3.5/3.0	RB	3.0/6.0
YR 4/5	.110	bR	3.0/4.5	yR	4.0/4.0	Y	4.0/2.0		Y 5/7	.097	yR	3.0/1.5	rY	3.0/2.0	bR	3.0/2.0
G 3/4	.107	rB	3.5/4.5	rB	4.0/2.0	G	4.0/2.0		G 3/4	.093	Y	3.0/0.5	yG	3.0/1.5	B	3.0/1.5
P 3/6	.097	rB	3.0/4.5	rB	3.0/3.0	G	3.0/2.0		YR 4/5	.069	RY	2.0/1.0	yR	3.0/3.0	bR	2.0/2.5
R 3/7	.049	RB	1.5/7.0	bR	2.5/7.5	RY	2.0/1.5		R 3/7	.048	yR	2.0/1.5	R	2.0/2.0	bR	2.0/3.0
A 0/0	.030	RB	2.0/5.0	RB	2.0/9.0	A	0.0/0.0		A 0/0	.030	yR	1.0/0.5	RY	1.0/1.0	A	0.0/0.0

samples, with only two exceptions, are R or YR, one of the exceptions being bR (the b is an after-image component), the other, A, coincides with expectation. Similarly with the other illuminants: on white ground lighter samples have either the illuminant hue or a hue representing a mixture of daylight and illuminant hue if the illuminant is not fairly homogeneous, darker samples have the after-image hue and samples of intermediate reflectance either are achromatic or have hues of low saturation.[11] The hues are similar on gray ground but for the fact there are fewer after-image and more illuminant hues. On black ground we find practically no after-image hues, illuminant hues appearing almost exclusively.

The relation of the hues of samples in strongly chromatic illuminants to their daylight hues appears from Table 2. Some measure of constancy is present as some of the samples have at least one component of their daylight hue even in the homogeneous illuminant (yR). Daylight hue by itself is not sufficient for predictions of constancy. Whatever constancy is found is as much a function of reflectance of sample for the chromatic illuminant as the commonly supposed factor of hue. The *lightness* of a sample furnishes a better clue to its behavior under different conditions than does its hue. This explains why a sample whose daylight hue is different from that of the illuminant may have a more saturated illuminant hue than one whose daylight hue is like that of the illuminant —*e.g.*, pB 7/4 is a more saturated red in red illumination on black ground than R 3/7 owing to its higher lightness; similarly R 7/6 is a more saturated green in green illumination on black ground with a value of 4.0 than G 3/4 with a value of 2.0. A 10/0 always has the highest saturation in all illuminants because of its highest lightness. Samples having the hue of the illuminant will retain this hue provided they are of *high* reflectance in the illuminant. Samples having hues complementary to the illuminant will remain complementary in an illuminant provided their reflectance is *low* in it. Achromatic samples will remain achromatic in chromatic illuminants

[11] These are the samples most difficult to predict at samples with small saturations are liable to change, shifting from illuminant to after-image hue and back.

only if their reflectance happens to be near 'adaptation reflectance,' otherwise they will take either the illuminant or after-image hues. Color constancy is thus seen to occur for different reasons which can all be subsumed under the general principle of color conversion which we shall state after discussing the relations of lightness and saturation.

From Table 2 it is seen that the relations of lightness and saturation are different on the three backgrounds and seem anomalous at first sight. On white and gray backgrounds, decrease in lightness is correlated first with decreasing and later with increasing saturation, while on the black background lightness and saturation are positively correlated throughout. Both of these cases are in turn at variance with the classical account of the saturation-intensity relation according to which maximum saturation is found with stimulus intensities of *intermediate* value.[12] How explain both our own findings and the facts which have been taken for granted in the literature and in textbooks for years? The answer lies in the relation of sample reflectance to 'adaptation reflectance' and the dependence of the latter on background reflectance. To these we now turn in our theory of conversion.

The theory of conversion makes use of the experimental facts so far discussed and but one assumption in addition. We assume that in every viewing situation there is established a level of adaptation corresponding to a certain lightness level such that samples having reflectances yielding this lightness appear either achromatic or a desaturated purple. Samples having reflectances higher than 'adaptation reflectance' will take the hue of the illuminant, samples below adaptation reflectance take the after-image hue and samples near adaptation reflectance are either achromatic or desaturated.[13] The

[12] F. Hillebrand, Purkinjesches Phänomenon und Eigenhelligkeit, *Z. f. Psychol. u. Physiol. d. Sinnesorgane*, Abt. 11, 1920, **51**, 46–95.

[13] This theory has developed out of the work at Bryn Mawr and from previous work of the senior author with D. B. Judd. The fundamental facts appeared in a study by Helson and Judd, A study in photopic adaptation, this *Journal*, 1932, **15**, 380–398, and in other studies referred to above by Judd and Helson. A mathematical formulation of the theory by Judd stated in the next paragraph in colorimetric terms will appear in *J.O.S.A.* A preliminary account by Judd appears in: Hue, saturation and lightness of surface colors with chromatic illumination, *ibid.*, 1939, **29**, 260.

present work with selective samples proves the validity of this formulation as well as the earlier work with non-selective samples.

The concept of adaptation level must not be regarded too simply. Only for non-selective samples are hue, saturation, and lightness simply a function of the reflectance of the sample (for the chromatic illuminants) relative to adaptation reflectance. For selective samples adaptation reflectance supplies a basic value in the definition of achromatic point on the triangular mixture diagram with reference to which hue and saturation are indicated. Adaptation reflectance has been shown to be a function of the reflectances of all samples in the field of vision, predominant weight being given to the background reflectance. The achromatic point, which we must determine anew for each sample of different spectral reflectance, depends chiefly upon the spectral character of the illuminant, upon the relation between adaptation reflectance and reflectance of the sample for the particular illuminant, and upon general level of illumination, to mention only the most important factors. Saturation may be regarded as a function of the distance of the sample point in the trilinear diagram from the achromatic point. Hue may be regarded as a function of the direction of the vector drawn from the achromatic point to the sample point.

The principle of conversion has related hue and saturation with lightness and we must now answer the question: "What determines the lightness of samples in various illuminants?" Reports in Table 2 bring out the almost perfect correlation existing between reflectance of samples for an illuminant and lightness. With changes of reflectance of a sample in different illuminants go changes in its lightness. Position of a sample in the reflectance column determines its position in the lightness column within the limits of experimental error. If lightness is a function of reflectance for an illuminant and as reflectance does not depend upon background, we should expect lightness values to remain practically constant with

changing backgrounds.[14] Average lightness values for all illuminants and intensities are 4.62 for white background, and 4.85 for both gray and black backgrounds (Table 3). We have here the basis for the well-known 'brightness-constancy.'

Influence of Background on Hue, Lightness, and Saturation

That hue and saturation depend as much upon background lightness as upon composition of illuminant and spectral reflectance of sample appears from Table 2. For identical samples have different hues and saturations when placed on different grounds, *e.g.*, P 3/6 in red illumination is gB 3.5/1.0 on white ground, Y 3.0/0.7 on gray ground, and bR 2.0/2.0 on black ground; similarly, G 7/7 in blue illumination is gB 7.0/4.5 on white ground, B 7.0/6.0 on gray ground and B 7.5/5.0 on black ground. The hue of any sample is thus a function of the relation of its reflectance to adaptation reflectance, which depends largely on background. Owing to the close correlation between reflectance and lightness we may say, though with less accuracy, that hue is a function of position in the lightness-continuum taken with respect to adaptation lightness. As Helson[15] has previously said in writing of the effect of background on hue: "By its preponderating influence on adaptation level, the background can determine whether objects shall be tinged with illuminant hue, the contrasting hue, or shall lose in saturation to the point of achromaticity."

The effect of background on saturation is closely linked with its effect on hue. *On white ground dark samples are most saturated* (*in the complementary hue*) *while on black ground light samples are most saturated* (*in the hue of the illuminant*). This is explained by the dependence of hue and saturation on adaptation reflectance and the effect of background on adaptation level. Since some samples gain while others lose

[14] Due to lightness-contrast this statement is not, of course, strictly true if identical samples are compared on backgrounds of different reflectances simultaneously. We shall return to this point again below.

[15] *Cf.* footnote 10, p. 410.

in saturation on any given ground it is interesting to determine if there is a net gain on one ground over the others. From Table 3 it is seen that samples on black ground are on the average 1.14 saturation steps higher than on white ground and the average for gray is 0.73 step higher than on white ground. Higher saturations are found on the darker grounds in all illuminants, as Table 3 proves, with only one exception. The increase in saturation on darker grounds is due to increase in number of illuminant hues and in their saturations at the expense of the complementary hues.

Turning now to background and lightness we find that this dimension is least affected by background. The average values in Table 3 are practically identical for the three back-

TABLE 3

AVERAGE SATURATIONS AND LIGHTNESSES OF SAMPLES ON A 10/0, A 5/0 AND A 0/0 BACKGROUNDS IN FOUR ILLUMINANTS

Ill.	Saturation				Lightness			
	A 10/0	A 5/0	A 0/0	Av.	A 10/0	A 5/0	A 0/0	Av.
R	2.84	3.77	5.08	3.90	4.39	4.56	4.52	4.49
Y	4.20	4.58	4.59	4.46	4.97	5.10	5.42	5.16
G	3.53	4.10	3.92	3.85	4.65	5.01	4.81	4.49
B	2.56	3.62	4.13	3.44	4.44	4.70	4.64	4.59
Av.	3.29	4.02	4.43		4.62	4.85	4.85	

grounds, a fact which seems to contradict the well-known facts of lightness contrast. When *two* identical samples are compared simultaneously on white and black grounds the higher lightness of the one on black ground is easily seen. But when *many* samples are viewed on a half-white-half-black background the average lightness values will not be higher on the darker portion as Ss raise all lightness values for the samples on the lighter side! Under the conditions used in these experiments, that is, with samples on either white, gray, or black ground and with the lightness-range prescribed, Ss tended to assign the maximum value of 10 to the lightest sample in the field and ordered the remaining samples accordingly. In this

way the lightness-inducing effect of darker grounds is lost in the reports and there results an appearance of even greater lightness constancy than is actually present.

But changes in intensity of illumination as well as changes in reflectance of background lighten or darken all samples simultaneously and the question arises: Do changes in general intensity affect lightness and saturation as much as changes in reflectance of background? Increase in general intensity from 1.4 f.c. to 135 f.c. increased average lightness from 4.50 to 4.96. Thus 96-fold change in intensity of illumination caused a change of only about one-half step in lightness so that lightness constancy is here about the same as with changes in reflectance of background. The case is different with saturation. This range of illumination intensities caused a change of only 0.75 saturation steps, on the average, whereas change in reflectance of background from 0.80 to 0.03 (about 27-fold change) resulted in an increase of 1.14 steps in saturation. If changes in general illumination were as effective as changes in reflectance of background, the 96-fold change in general intensity should have caused an increase of 4.9 steps instead of 0.75. Background lightness is thus seen to be about six times more effective than general intensity in effecting changes in saturation.

Determination of Adaptation Level

The determination of adaptation level is important for predictions involving color conversion. In assuming reflectance of background exercises preponderant influence in determining adaptation reflectance and its associated lightness, which we may call 'adaptation lightness' (L'), Judd,[16] and following him, Helson,[17] have taken adaptation reflectance to be the weighted geometric mean of background and sample reflectances according to the following formula:

$$A' \equiv (\bar{A} \cdot A_0^3)^{1/4}, \qquad (1)$$

where A' is adaptation reflectance, A_0 is reflectance of back-

[16] D. B. Judd, Hue, saturation and lightness of surface colors, *J.O.S.A.*, in press.

[17] *Cf.* footnote 1, p. 453.

ground, and $\bar{A}$ is the logarithmic average reflectance of all samples in the field of view. To predict lightness values in conformity with the 0–10 scale used in these experiments when reflectances range from 0.03 for darkest samples and background to 0.80 for lightest, Judd proposed the following formula by which reflectances transform into lightnesses:

$$L \equiv \frac{1.03(10A - 0.3)(A' + 1)}{A' + A}, \qquad (2)$$

where A' is defined by formula (1) now and A is the apparent reflectance of the sample whose lightness is desired. To obtain adaptation lightness we substitute A' for A in formula (2) since we desire the lightness corresponding to adaptation reflectance or L'.

Formula (2) gave results in excellent agreement with the average lightness reported achromatic in a former study by Helson [18] with non-selective samples when $4/5A'$ was used in (2) in place of A'; but when applied even with this correction to the data of Table 2, it yielded values too high for adaptation lightness. Whereas adaptation lightness with only non-selective samples in the field was 7.1, 4.8, and 0.4 on the white, gray and black grounds, we find it to be 4.50, 3.81 and 0.00 in the present work with selective samples. Before making the changes in formula (1) necessary to make (2) satisfactory let us consider some of the differences between the former study and this one which account for the lower adaptation level we now find.

The chief differences between conditions in the present study and the earlier are to be found in the greater difficulty in judging saturations and hues of selective samples. Here we find greater individual differences among Ss and inversions in individual reports are more frequent. While plots of saturation as a function of lightness yield curves having the same general form found for non-selective samples, the scatter is much greater owing to longer fixation of samples and other factors which need to be correlated with measureable varia-

[18] *Ibid.*

bles. At the moment we can only 'allow' for them by introducing constants into the formulæ to accord with the experimental data.

There are, however, certain factors in the present conditions which are measureable and account, in part, for the lower adaptation lightness. Chief of these is the higher average reflectance of the samples in the present study as compared with that in the previous one. Whereas the logarithmic average reflectance of the non-selective samples was 0.14, here it is 0.22 for all samples in all illuminants. For a given reflectance of background the presence of a larger number of samples of higher reflectance will produce *more illuminant hues* and the situation is *as if* the reflectance of the background had been lowered, without this being the case, however. This means that the value of A' must be reduced if we are to derive values of L' in agreement with the findings of the present study.

We therefore change Judd's formula (1) as follows: we take the logarithmic average of $\bar{A}$ and A_0, giving unit weight to each, and then multiply this by the reducing factor 0.32. This gives:

$$A'' \equiv 0.32(\bar{A} \cdot A_0)^{1/2}. \qquad (3)$$

The use of A'' in formula (2) for L' yields adaptation lightnesses in good agreement with average observed adaptation lightness as shown in Table 4.[19] Owing to the inversions in

TABLE 4

AVERAGE LIGHTNESS OF SAMPLES REPORTED ACHROMATIC IN ALL ILLUMINANTS AND ADAPTATION LIGHTNESS AS CALCULATED FOR EACH BACKGROUND

Bckgd.	Observed L'	Calculated L' ($A'' = 0.32(\bar{A} \cdot A_0)^{\frac{1}{2}}$)
White	4.50	4.51
Gray	3.81	3.25
Black	0.00	0.00

the data average lightness for achromaticity was taken as the mean of the lightnesses of the two lowest contiguous satura-

[19] Calculations for L' are based on the following values derived from Table 2: logarithmic average reflectance of all samples in all illuminants, $\bar{A}$, is 0.22; A_0 is 0.80 for white ground, 0.24 for gray ground, and 0.03 for black ground. To solve for L' in formula (2) set $A = A''$.

tions on white and gray grounds and as zero on black ground because A o/o is achromatic on A o/o ground in every illuminant.[20]

The *A''* formulation for adaptation reflectance with its associated adaptation lightness still gives predominant weight to background reflectance in determining adaptation level but allows for special factors entering into these experiments.

Tolerance to Chromatic Illumination

The extent to which samples may be subjected to chromatic illuminants while retaining more or less of their daylight hues can be regarded as a measure of 'constancy.' Through the use of two apertures, one for filtered light and the other for lamp light of 2848° K., it was possible to mix varying amounts of these lights. Owing to the differences in transmission of our filters the addition of any given amount of lamp light to the filtered light resulted in different *percentages* of mixture: thus the use of 0.3 f.c. lamp light with the various filters gives the following percentages of lamp light in the mixtures when the illumination was 72 f.c. on sample plane without any filters: 8 percent in the yR, 1 percent in the rY, 5 percent in the yG, and 68 percent in the rB illuminations. The higher percentages in the yR and rB compensate for the greater chromaticity at lower intensities of wave-lengths at the ends of the spectrum. Thus in spite of the greater percentages of lamp light in the yR and rB illuminations the effects on hue are very much like those for the rY and yG.

Comparison of the hues reported in Table 2 with those given in Table 5 shows the effect of adding lamp light to the highly chromatic illuminants. A few examples will illustrate the general tendency: Y 7/8 which is RY in homogeneous yR illumination on white background is Y with the addition of the lamp light; G 3/4 under the same conditions changes from GB to bG, the added heterogeneous light reducing the B of the after-image hue; B 7/4 in yG illumination on white ground

[20] Adaptation reflectance turns out to be 0.026 on black ground but owing to formula (2) values less than 0.03 for *A''* give negative values of *L'*. The small negative value in this case we take to be zero.

is yG but with the added lamp light it becomes gB, the yellow component of the illuminant disappearing while the green is reduced and the daylight hue becomes most prominent. The return to 'normal' is not a simple matter. Lightness and background reflectance come in to complicate matters. Thus G 3/4 in yG illumination on white ground is B, while G 7/7 is yG, the lighter sample holding the illuminant hue while the darker keeps the after-image hue.

The effect of background on constancy can be seen by following any one sample on the different backgrounds but in order to get a general idea of the facts let us deal in totals. Inspection of reports on black ground reveals fewer daylight hues than on white ground. Of the 180 reports on selective samples found in Table 5, 22 fail to contain at least one hue-component characteristic of the samples in daylight and of these 22, 12 occur on black ground as against 10 for both white and gray grounds. The fact that A o/o is A o/o with black ground in every illuminant is not an exception to our rule for it is the exception that proves the rule: A o/o being very near to adaptation reflectance on black ground, we expect it to be reported A o/o in homogeneous as well as mixed illuminations.

We come now to a consideration of the non-selective samples which we have left out of the discussion thus far. The statement is often found in the literature that "brightness constancy is greater than color constancy." This statement is extremely ambiguous as it stands for if taken in one way it is correct while taken in another it is incorrect. It is true that the lightness of both selective and non-selective samples undergoes less variation than hue with changes in composition and intensity of illumination and reflectance of background. It is not true that non-selective samples keep their achromaticity in chromatic illuminations better than selective samples do their hues. Exactly the reverse is the case because selective samples through their differential absorption of the components in the illuminant are able to emphasize certain wave-lengths and thereby their 'own hues' whereas non-selective samples, returning the energy of the illuminant indis-

TABLE 5

HUE, LIGHTNESS, AND SATURATION AND APPARENT REFLECTANCE (A) OF 18 MUNSELL SAMPLES (M) IN FOUR CHROMATIC ILLUMINANTS MIXED WITH 0.30 F.C. LAMP LIGHT (2848° K.)

M	Yellowish-Red Illumination						Reddish-Yellow Illumination					
	A	White		Gray		Black	A	White		Gray		Black
		H	L/S	H	L/S	H	L/S		H	L/S	H	L/S	H	L/S
R 3/7	.263	bR	5.0/3.0	yR	5.0/6.5	R	5.5/7.0	.151	bR	4.5/7.5	bR	4.5/7.0	yR	5.0/7.0
R 5/10	.671	yR	7.0/6.0	yR	7.0/7.0	yR	7.0/9.0	.402	yR	7.0/6.0	yR	7.0/8.0	yR	7.0/7.5
R 7/6	.759	RY	8.0/4.5	RY	7.5/7.5	yR	7.0/8.5	.616	yR	7.5/4.0	rY	7.5/7.0	rY	7.5/8.0
YR 4/5	.288	rY	5.0/2.0	rY	5.0/5.0	R	6.0/7.0	.245	yR	5.0/2.0	rY	5.0/3.0	Y	6.0/5.0
Y 5/7	.271	G	5.0/0.5	gY	5.0/3.0	yR	5.0/6.0	.288	A	5.0/0.0	yG	5.0/3.5	YG	5.0/5.0
Y 7/8	.573	Y	7.0/4.0	rY	6.5/6.0	yR	7.0/8.0	.588	yG	8.0/3.5	gY	7.0/6.0	Y	8.0/7.0
G 3/4	.074	bG	2.0/3.0	bG	2.0/1.0	bR	4.0/2.0	.079	B	2.0/3.0	B	3.0/3.0	yG	4.0/4.0
G 5/7	.100	bG	3.5/7.0	G	4.0/8.0	bG	5.0/5.5	.172	G	5.0/7.0	G	6.0/8.0	G	5.0/8.0
G 7/7	.343	bG	6.0/4.0	bG	6.0/5.0	bR	6.5/5.0	.429	yG	7.0/6.0	yG	6.5/5.5	yG	7.5/6.0
B 3/5	.067	gB	2.0/4.0	gB	2.0/4.0	rB	3.0/3.0	.073	B	2.0/3.5	B	2.5/2.0	G	3.0/3.0
B 5/6	.168	gB	3.5/3.0	B	5.0/3.0	RB	5.0/4.0	.185	BG	4.5/1.5	bG	5.0/4.0	yG	5.5/4.0
B 7/4	.391	gB	7.0/1.5	B	7.5/1.0	bR	7.0/5.0	.446	bG	7.0/5.0	yG	7.0/5.0	gY	7.5/5.0
P 3/6	.150	rB	2.5/2.0	bR	3.5/4.0	bR	4.0/6.0	.119	RB	4.0/5.0	bR	4.0/3.0	yR	5.0/4.5
P 5/6	.351	RB	6.0/2.5	bR	5.0/5.0	yR	7.0/7.0	.273	yR	5.0/2.5	bR	6.0/3.5	rY	6.0/4.0
pB 7/4	.429	rB	6.0/1.0	RB	7.5/4.0	bR	6.5/6.0	.405	gB	7.0/0.5	gY	6.5/2.0	gY	6.0/5.0
A 10/0	.800	R	9.0/0.5	R	8.0/3.0	R	7.0/7.5	.800	Y	9.0/3.0	Y	8.0/7.0	Y	8.0/8.0
A 5/0	.240	gB	4.0/0.5	R	4.5/1.0	bR	5.0/6.0	.240	RB	5.0/0.5	Y	5.0/0.5	Y	5.0/4.0
A 0/0	.030	BG	0.5/3.0	GB	1.0/2.0	A	0.0/0.0	.030	rD	1.0/4.0	B	0.5/0.5	A	0.0/0.0

TABLE 5—*Continued*

M	Yellowish-Green Illumination						Reddish-Blue Illumination					
	A	White		Gray		Black	*A*	White		Gray		Black
		H	*L/S*	*H*	*L/S*	*H* *L/S*		*H*	*L/S*	*H*	*L/S*	*H* *L/S*
R 3/7	.054	RB	3.5/7.5	bR	3.0/6.0	yR 3.0/4.0	.072	R	3.0/3.0	R	3.0/7.0	R 4.0/6.0
R 5/10	.127	bR	4.0/7.5	bR	4.5/6.5	yR 5.0/6.0	.180	yR	4.0/7.0	bR	5.0/7.0	bR 5.5/7.0
R 7/6	.373	bR	7.0/3.5	rY	7.0/3.0	gY 7.0/4.5	.413	R	6.0/5.5	bR	7.0/5.5	bR 7.0/5.0
YR 4/5	.118	rY	4.5/4.0	rY	4.5/6.0	Y 5.0/4.0	.119	yR	4.0/6.0	RY	4.5/6.0	yR 4.5/6.0
Y 5/7	.225	rY	5.0/4.0	Y	5.0/4.5	gY 5.0/4.0	.166	rY	4.0/5.0	rY	4.0/6.0	RY 5.0/5.0
Y 7/8	.462	Y	7.0/5.0	Y	7.0/6.5	gY 7.0/6.0	.316	rY	6.0/7.0	rY	6.0/8.0	RY 6.0/6.0
G 3/4	.105	B	3.0/2.0	bG	4.0/0.5	G 4.0/5.0	.093	G	2.5/0.5	bG	3.0/4.0	B 4.0/1.0
G 5/7	.364	bG	6.5/6.0	G	6.0/7.0	G 6.0/7.5	.262	G	6.0/7.0	G	6.0/6.0	bG 6.0/7.0
G 7/7	.586	yG	8.0/4.0	yG	7.5/4.5	yG 7.0/6.0	.494	yG	7.0/5.0	bG	7.0/3.0	gB 8.0/3.0
B 3/5	.129	B	4.0/5.5	B	3.5/4.0	bG 4.0/7.0	.166	GB	5.0/6.0	gB	5.0/7.5	B 5.0/8.0
B 5/6	.138	gB	5.5/5.0	bG	6.5/7.0	G 6.0/6.5	.377	gB	6.0/7.0	gB	6.5/7.5	B 6.0/6.5
B 7/4	.573	gB	8.0/3.0	yG	8.0/3.5	yG 7.0/6.5	.584	B	8.0/5.0	B	8.0/4.0	B 8.0/5.0
P 3/6	.098	RB	3.5/6.0	rB	3.5/2.0	yG 3.0/0.5	.135	rB	4.0/5.0	rB	4.5/6.0	RB 4.5/8.0
P 5/6	.238	RY	5.0/0.5	bG	5.5/0.5	yG 5.0/4.0	.288	rB	5.0/5.0	rB	6.0/6.0	rB 6.0/5.5
pB 7/4	.481	gB	7.0/3.0	G	7.0/3.5	yG 7.0/5.5	.531	B	7.5/5.0	B	7.0/5.5	B 7.0/6.0
A 10/0	.800	YG	9.5/1.0	gY	8.0/3.0	gY 8.0/5.0	.800	A	10.0/0.0	rB	9.5/0.5	RB 8.0/4.0
A 5/0	.240	RB	5.0/1.0	A	5.0/0.0	yG 5.0/4.5	.240	A	5.0/0.0	A	5.0/0.0	rB 5.0/5.0
A 0/0	.030	rB	1.0/2.0	RB	1.0/0.5	A 0.0/0.0	.030	A	0.0/0.0	A	0.0/0.0	A 0.0/0.0

criminately, do not retain their achromaticity and are therefore more subject to conversion.[21] The non-selective samples in Table 5 are seen to lag behind the selective samples in returning to daylight hue. This is confirmed by adding 4.5 f.c. lamp light to the chromatic illuminants. While this amount of lamp light sufficed to bring back almost 100 percent of the selective samples to their daylight hues it was much less effective with the non-selective samples, 36 percent failing to become achromatic in it.

The quantitative determination of tolerance to chromatic illuminants, hence of constancy, is beset with the difficulty of drawing the line at which samples may be said to have their 'normal' hues. We have seen that with as little as 0.3 f.c. lamp light some return to daylight hues is noticeable (Table 5). If we take as our standard of tolerance the complete return of all samples to their daylight hues the amount of tolerance is enormously reduced for the lamp light required for this purpose is greater than the amount of chromatic in most cases.

For practically complete return of selective samples to daylight hues 4.5 f.c. lamp light was necessary. Owing to the differences in transmission of the four filters this gave the following percentages lamp light in the chromatic: 120 percent for the yR, 17 percent for the rY, 167 percent for the yG, and 1002 percent for the rB. Obviously the amount of lamp light admixture varies greatly for the different hues and brings out the fact that tolerance to chromatic components in an illuminant is least for the hues which have the greatest saturations.[22] With the exception of the yG the reciprocals of these percentages follow the order of intensities necessary for maximum saturation, violet and red requiring minimal intensities and yellow and green maximal. Had we used a graded series

[21] In line with this finding we recommend the use of non-selective samples to detect chromatic components in weakly chromatic illuminants. Non-selective samples furnish a more severe test of the achromaticity of an illuminant than do selective samples which have far greater constancy.

[22] In this connection the results for after-desaturation should be noted. *Cf.* also footnote 9.

of admixtures of lamp light to each of the chromatic illuminants the yG illuminant would not have emerged as an exception in all probability. On the other hand, the value for red would have been larger had our yR illuminant plotted in a longer-wave region of the spectrum (say 660 mμ) instead of at 630 mμ which is more yellowish-red. This would bring our results into perfect alignment with the facts of chromatic valence and the greater fatigability of the red and violet as opposed to the yellow and green hues.[23]

Some Theoretical Conclusions

The methods employed in the present series of studies have been characterized by stimulation of the *whole retina* with homogeneous and near-homogeneous illuminants. In this way we have duplicated the classic use of stimuli of known composition but we have gone farther in that we have varied the conditions of stimulation by enlarging the retinal area subject to these stimuli and by introducing differential intensities of stimulation between retinal regions. Our results prove the importance of field conditions in determining the color of local regions. This determination is wider than can be inferred from the classical work and theories based on it. Backgrounds differing only in lightness have been found to exert not only the classical lightness contrast but hue and saturation effects of even greater importance. We would expect, and confidently predict, that experiments employing small spectrally homogeneous patches of light will yield different results when background (surround in the retina) is changed from the classic dark field to light or chromatic field.

In a previous study in this series it was emphasized that underneath the *various color phenomena* a *single mechanism* must be functioning because effects formerly attributed to entirely different processes could be seen simultaneously if

[23] It should be pointed out that attitude of the S has a good deal to do with the amount of tolerance which will be found in any case. An S who is set to discover as many daylight hues as possible will seem to have greater tolerance for chromatic illuminants. Color workers, to whom small color differences are important, will show less tolerance than Ss with the other attitude. Quantitative measurements of tolerances will eventually have to provide for such differences in attitude.

the right conditions were provided. The principle of conversion states both the phenomenological or color facts and at the same time indicates the mechanism by which they are to be explained. Let us assume, as Troland has done in his presentation of the trichromatic theory,[24] that each intensity and composition of light beam reaching the eyes causes reaction in each of three retinal substances. Equal rates of the three result in achromaticity and hence determine adaptation level. A certain hue would correspond to a higher rate and the after-image hue to a lower rate. Actually both rate of reaction of processes in the retina and ratios of rates in different parts of the retina must be taken into consideration because higher saturations are found both with higher general intensities of illumination and with larger differential intensities within the visual field. High rates of reaction giving unit ratios among the three substances result in high adaptation level (white) while low unit ratios yield a low adaptation level (dark gray or black).

It is of minor importance whether we envisage these facts in terms of one color theory or another. On the basis of the Hering theory it is necessary to think of a given hue, which may arise as an illuminant hue or as a complementary hue, as being due in one case to ratios greater than unity and in the other case to ratios less than unity. This is awkward and so a trichromatic formulation seems at the moment preferable. Assuming now that hue is a resultant of the excitations of three color-cones or substances in the retina which respond in varying amount to various wave-lengths in the usual manner, we would expect these substances to mediate hues according to the rate at which they produce color materials through decomposition or some other physico-chemical change. Since rate of reaction depends on the concentration of reacting materials, as the reaction proceeds the concentration decreases and so does the rate. The faster each of these substances decomposes the greater its contribution to the resultant color but as a result less and less materials are available (up to a

[24] L. T. Troland, *The Principles of Psychophysiology*, 1930, Vol. 2, Sensation, 189 ff., especially 191–193.

certain point without replenishment) for further decomposition as stimulation continues. If one substance is used up more than the others, the latter make relatively more contribution to the seen color. With, for example, predominantly long-wave stimulation, one of the substances, the R-substance, will be stimulated more by the lighter parts of the field than the G- and V-substances and will contribute more to the resultant hue. But owing to the exhaustion of the R-substance darker parts of the field which do not send enough energy to the eye to maintain a higher rate of reaction than that of the G- and V-processes will appear not in R-hue but in GV-hue. The G- and V-substances having decomposed less than the R-substance are relatively more sensitive. The lessened sensitivity of the R-substance is therefore manifested either in lowered saturation of R-hues or in the appearance of after-image hues.

The relation of reactivity to intensity, both absolute and differential, explains why increase in intensity of the whole field results in more illuminant hues in lighter parts of the field. The background governs both absolute and differential intensities very largely and thus exercises its effects. Shifting adaptation level is explained by reference to establishment of unit ratios at various intensity levels. White and black may be regarded most simply as high and low unit ratios. For vision in daylight with its rich variety of colors we assume an adaptation level corresponding to unit ratios between the R-, G-, and V-processes. Added stimulation of one or the other of the three substances from local parts of the field gives rise to local colors. The induction of black in dark parts of the field is explained by the lower unit ratios in these regions.

No claim is made for the complete adequacy of the trichromatic theory or any other theory of color vision. A formulation is necessary in physiological terms which brings all parts of the visual field into intimate relations and takes into account the basic fact, which we have found in all our work, that the eye is a mechanism which responds in very important ways to differential stimulation. The theory of conversion recognizes these things and yet does not depend

upon any particular theory of color vision since it springs directly from observation. Our application of the trichromatic theory to the phenomena of conversion has been merely for the purpose of illustrating how one of the classical theories can be made to explain new facts.

Summary

Experiments involving stimulation of the whole retina with homogeneous and near-homogeneous illuminants and differential intensities in various parts of the field show that certain hitherto neglected factors profoundly influence hue, lightness, and saturation in predictable ways. We can summarize briefly:

1. Hue, lightness, and saturation depend not only upon the composition and intensity of light from an object but fully as much upon the reflectance of background and other objects in the field of vision.

2. Through its effect on adaptation level the background may induce the illuminant hue or the after-image hue on the sample, depending upon the reflectance of the sample. Effects of backgrounds differing only in lightness are thus seen to be wider than classic laws imply, for hue and saturation are affected by reflectance of background as well as lightness.

3. Selective samples have greater constancy in chromatic illuminants than non-selective samples as the latter tend to be tinged more easily either with the hue of the illuminant or its complementary while the former tend to keep their daylight hue if their dominant wave-length is present even as a minor component in the illuminant.

4. The hue of selective samples in strongly chromatic illuminants which are not homogeneous tends to be a mixture of daylight hue and the hue resulting from conversion in homogeneous illuminants.

5. The effects of chromatic illuminants depend upon the distances of the illuminant points in the color mixture diagram from the white point. Illuminants having hues characteristic of the ends of the spectrum (red and blue-violet) give the greatest chromatic effects and after-effects.

6. Constancy of object color or tolerance to chromatic illumination is not predictable from the hues of sample and illuminant alone. It depends chiefly upon reflectance of sample and background, dominant wave-length of the illuminant, and attitude of the observer.

7. A revised formula was found necessary for adaptation reflectance to meet the conditions in the present study. It yields values of adaptation lightness in good agreement with observation.

8. The bearing of certain results in this study on previous work and theories has been pointed out and the importance of the eye as an organ that responds primarily to differential stimulation receives new confirmation.

(Manuscript received July 5, 1939)

A CONVENIENT AND PRACTICAL MEANS FOR STUDYING LIGHT AND COLOR MINIMA IN ANY PART OF THE RETINA

BY C. E. FERREE AND G. RAND

From the Research Laboratory of Physiological Optics, Baltimore, Maryland

The special purpose of this paper is to describe an attachment to the Ferree-Rand Perimeter for determining light and color minima in any part of the field of vision. The description is accompanied by a discussion showing the need and usefulness of the attachment. It has also been our wish in the preparation of the paper to contribute to a broader and better general understanding of the importance and practical value of studying the sensitivity of the retina to light and color. The most direct and important application of this study is at the present time, and perhaps always will be, in the medical field.

Three methods have been used to study light and color sensitivity: the light and color minima and light and color differences; the limits of sensitivity (the perimetric method); and the light minimum for the discrimination of detail.

1. *The Light and Color Minima and Light and Color Differences.*—These methods are customarily used to measure light and color sensitivity in central vision. They have not been used for practical work in peripheral vision because of the lack of a suitable and feasible means of presenting at all points in the field of vision a stimulus whose intensity can be varied. The results obtained may be expressed in various ways. In the use of spectrum light in scientific work we have been accustomed to make the specification in radiometric or energy units (watts or ergs). In the use of white light and colored light obtained from filters and pigment surfaces this is not feasible because of the difficulty of excluding the invisible radiations. In these cases the specification has been made

in photometric terms, *e.g.*, brightness of test surface (millilamberts, etc.), illumination of test surface (foot-candles, meter-candles, etc.), amount of light entering the eye (lumens per total area or unit area of test surface).

2. *The Limits of Sensitivity (the Perimetric Method).*—In this method, the one customarily used to study sensitivity in peripheral vision, the points farthest from the center of the field are determined at which a given stimulus is visible. In the minimum or threshold method, on the other hand, the intensity is determined at which a given stimulus is visible at a given point in the visual field. The latter method is the older and without doubt the logical procedure; the former may be regarded as an indirect and substitute procedure, in many respects a poor one.

Among the shortcomings of the method of limits, the following may be noted.

(*a*) Comparatively few points of the retina can be tested and the location of these points is determined by the visibility of the stimuli used. As the method is used at present, the only means available for producing changes in the limits of sensitivity are changes in the size of the stimulus and in its color and brightness relation to the background (1). With neither of these is it feasible or conveniently possible to make a very wide survey of the sensitivity of the retina. In using the method of limits the examiner has therefore to be content with a very meager knowledge of the functional condition of the retina. Too often he is not able to test sensitivity where the test is most needed. Indeed, handicapped as we have been by having only the method of limits at our disposal, we do not even know with certainty what are the most important points at which the test should be made, for example, for detecting different pathologic conditions and for differentiating between them. Even a brief consideration of the structural and functional differences from center to periphery of the retina in the various meridians and of the variety of disturbances that may occur will disclose the gross inadequacy of the method of limits as the sole test procedure and the very great advantage of adding to it the method of minima where

more detailed knowledge is needed. It is not difficult to understand that very little knowledge can be obtained of the sensitivity of the retina by mapping the field with stimuli of one, two or three sizes, particularly in case of the form field—little more, so to speak, than could be had of the details of a large country by travelling along its coast line or passing around its boundaries.

(*b*) Even with the improvements that have been made in perimetry, the method of limits is at best a rough means for detecting loss of sensitivity and one that varies a great deal in its delicacy for this purpose in different parts of the retina. This is due to the amount of difference there is in sensitivity at points separated as little as 1 deg. in the field of vision. We select 1 deg. here because changes in the limits of less than 1 deg. are not ordinarily detected. The delicacy of the method, therefore, can be judged by the loss of sensitivity that is required to change the limits by 1 deg. This varies greatly, for example, from center to periphery, being in general comparatively small in the paracentral portion, larger in the mid region and very large towards the far periphery. It is very great indeed around the blind spot and ordinarily very large around the borders of scotomas. A very great loss in sensitivity would have to be experienced around the blind spot, for example, to displace its margin as much as 1 deg. It is just in these locations where the method of limits is most deficient that great delicacy and high sensitivity of method are needed in the most important application of the method, namely, the study of pathologic conditions. There is no limit however to the sensitivity of the method of minima other than is imposed by the means which are used to produce the changes of the intensity of the stimulus. There is no limit at all when the intensity of the stimulus can be changed in continuous series.

As showing the variation in the intensity of stimulus that is required to change the limits of sensitivity by 1 deg. in different parts of the field of vision, Table 1 has been prepared. In this table the change in the minimum per deg. of the visual field is shown for three spectrum colors—red (670 mμ), green

(522 mμ) and blue (468 mμ). These data are taken from a previously published study (**2**) of color minima carefully determined at near-lying points from center to periphery in two meridians, specified in terms of watts $\times 10^{-12}$ entering the eye. At any of these points it will be remembered that the functional condition of the retina can be tested by the method of minima within the limits of delicacy of a just noticeable change of sensation.

(*c*) The method of limits has in general from the beginning been characterized by inaccuracy and variability of result, a fault so serious as formerly to have rendered the method prac-

TABLE 1

DIFFERENCES IN THE INTENSITY OF THE STIMULUS WHICH GIVE LIMITS AT POINTS SEPARATED BY 1 DEG. FROM CENTER TO PERIPHERY OF THE VISUAL FIELD

Red (670 mμ)		Green (522 mμ)		Blue (468 mμ)	
Region Examined, 180 Deg. Temporal Meridian (Deg.)	Change in Color Minimum per Degree (watt $\times 10^{-12}$)	Region Examined, 180 Deg. Temporal Meridian (Deg.)	Change in Color Minimum per Degree (watt $\times 10^{-12}$)	Region Examined, 180 Deg. Temporal Meridian (Deg.)	Change in Color Minimum per Degree (watt $\times 10^{-12}$)
5–10	1.4	5–10	0.4	5–10	1.1
20–25	4.1	20–25	3.3	30–35	2.0
30–35	20.6	35–40	4.0	50–55	8.1
60–65	69.9	48–50	15.0	75–76	90.7
65–70	127.4	50–51	370.0	76–77	725.4
80–82	328.8	59–61	715.0	78–80	2135.9
85–87	2630.4	63–65	4720.0	83–85	2700.1
87–88	5425.2			87–88	3465.8

tically worthless, particularly in the applied fields. Studies of factors and the development of improved methods, controls and instruments have done much in the way of correction in later times, and the study of the visual field by this method has now become standard practice. But the method is too often not successful in practical work because accuracy of result can be secured only by an amount of care greater than many examiners are willing to give. A point in question is the determination of the limit with a moving stimulus. Acceptable accuracy of result cannot possibly be obtained in this way. A good control of fixation cannot be had with a moving stimulus and without a good control of fixation accuracy in

the determination of the limit is not possible. Further, accuracy in the determination can not be had without a proper control of the brightness of the preëxposure, particularly in case of the color limits. A moving stimulus may be used to make a rough determination of the limit, but the final determination should be made with a stationary stimulus and proper precautions as to fixation, preëxposure and exposure. In the method of minima, on the other hand, the exact point at which the result is to be obtained is known in advance, and the determinations, therefore, can and naturally are made with the proper conditions of preëxposure and exposure and with proper control of fixation. It is quite clear that inaccuracy and variability of result are inherent in the method of limits and can be avoided only by an amount of care and caution that many examiners are not willing to give.

Perhaps the most serious fault of the method of limits, as it is now used, is the examiner's inability to make the test where he wishes. Examples of the need of full freedom in this respect in the medical field are the early stages of detachment of the retina and during the course of reattachment, all disturbances in the central, paracentral and mid-peripheral portions of the retina, early stages in the disturbance in all parts of the retina and all cases of advance or recession of the disturbance. The complete adequacy of the threshold method in all such cases and the serious inadequacy of the method of limits should be clear to everyone. Examples of the need for freedom in other fields vary with the problem under study and are too numerous to mention here.

Recognizing the usefulness of both the method of limits and the method of minima, we have devised the attachment to the Ferree-Rand Perimeter (**3**) described in this paper. By means of this attachment both the light and color minima can be easily and conveniently determined at any point in the field and the method of limits can be extended to include the use of stimuli of different intensities in a continuously graded series.

3. *The Light Minimum for the Discrimination of Detail.*— Sensitivity to light may be measured directly or in terms of its relationship to the seeing of objects; that is, either the

light minimum may be determined or the minimum amount of light required to discriminate detail in objects. Both procedures have significant and practical value. However, if light sensitivity is measured directly, its relationship to the seeing of objects can not be quantitatively inferred; nor can the effect of differences in light sensitivity in different individuals or at different ages, of change in light sensitivity due to adaptation, or of disturbances due to pathologic conditions on the power to see objects be correctly inferred from determinations of the light minimum. The special usefulness of the direct measurement of these effects in the testing of fitness for various kinds of work and types of ocular performance is obvious. So far as we know, the method has not been sufficiently used in the study of eye diseases to show whether it has any very great differential or diagnostic significance. However, quite apart from this it is our belief that it should be included in every important program for the study of the characteristics of eye diseases. Where it is of advantage to do so, this method can also be used for the study of sensitivity to colored light and light tinged with color, or light having a dominant hue.

For the application of this method the Projector devised by us and manufactured by Bausch and Lomb Optical Company is admirably adapted. For a description of this Projector and a discussion of its use in determining the light minimum for the discrimination of detail, the reader is referred to "Testing fitness for night flying: visual acuity," *Archives of Ophthalmology*, 1938, **20**, 58–79. The work done with the instrument in relation to night flying at the School of Aviation Medicine, Randolph Field, may also be of interest (**4**).

An Attachment to the Ferree-Rand Perimeter for Determining Light and Color Minima

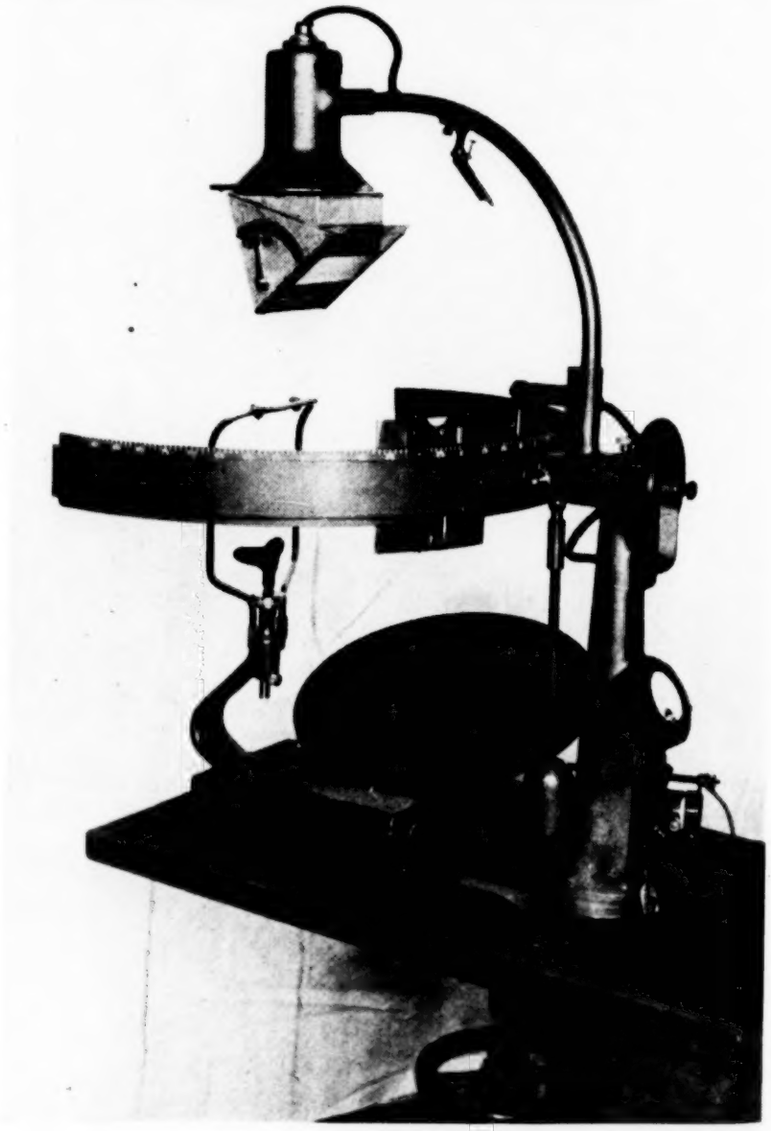
This attachment is in the form of a lamphouse so constructed that it can be easily substituted for the lamp and housing originally supplied with the perimeter. The purpose of the attachment is to provide a means of varying the intensity of illumination on the perimeter arc in continuous

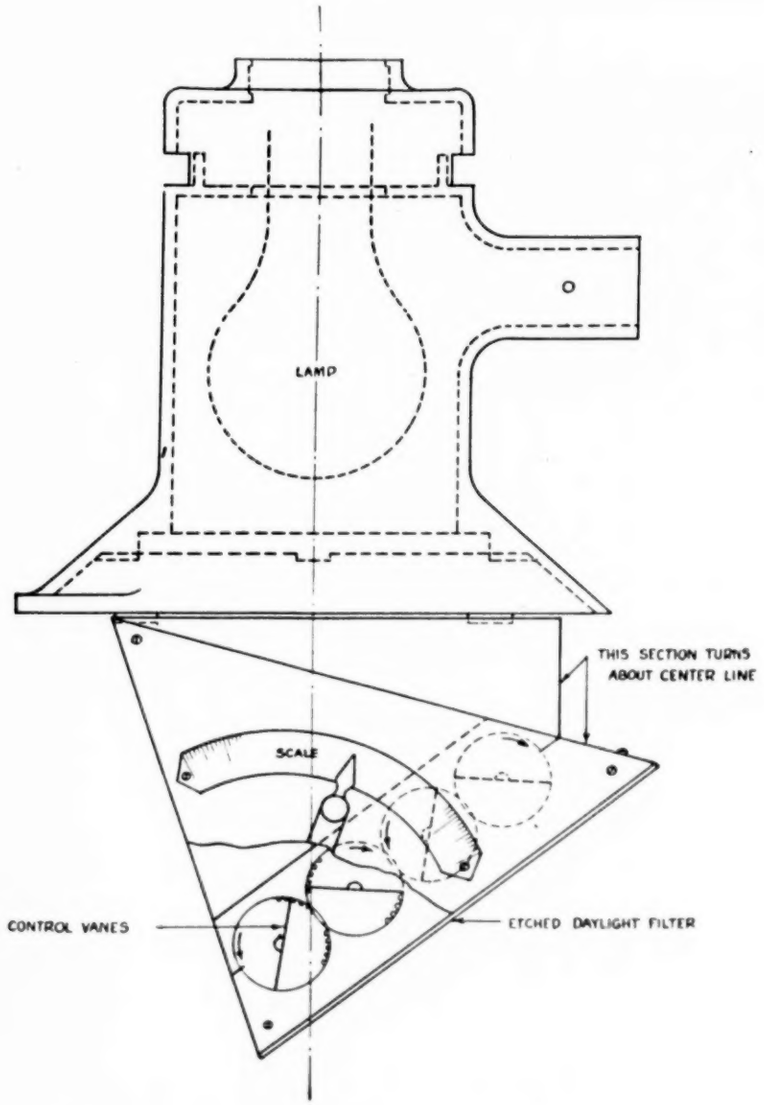
change from zero to full without altering the color or composition of the light or the size, shape or position of the illuminated area. The essential features of the attachment are a source of light of adequate intensity to give the range needed, a housing of suitable size and shape, a mechanical means or specially designed shutter for varying the intensity of the light, a diffusing plate or other diffusing means both to eliminate completely the shadows that would otherwise be produced by the shutter and to give an evenly distributed and well diffused illumination of the test object and its surroundings, and a filter to correct the artificial light to daylight quality.

A photograph of the attachment in position on the perimeter is shown in Fig. 1. At the right of the photograph is given an enlarged drawing of the attachment, side section elevation.

The shutter consists of four vanes which extend across the opening of the lamphouse in such relationship to each other that when their flat surfaces are parallel to the beam of light, the maximum amount of light passes through the opening and when they are rotated to a position at right angles to the beam, the light changes in continuous series from full intensity to zero. The shutter is so actuated as to cause the contiguous vanes to turn in opposite directions. This insures that the variation of intensity is not accompanied by a shift in the position of the illuminated area or of change in its size and shape as would be the case if the vanes all moved in the same direction. The vanes are painted flat black.

In various models of the shutter different means of actuating the vanes have been employed. In the model shown in Fig. 1, to each vane is fastened along its longitudinal axis a shaft one end of which passes to the outside of the housing. Mounted at the outer end of each of these shafts are small cogwheels or gears which mesh with each other. The cogwheels are turned by means of a small knob on the end of a pointer or indicator attached directly to the shaft of one of the vanes. When the cogwheels are turned through 90 deg. the vanes pass from the position of full light to complete ex-





Enlarged drawing of attachment, side section elevation.

FIG. 1. Ferree-Rand Perimeter with its attachment for determining light and color minima in any part of the visual field.

tinction and the pointer passes over a graduated scale which indicates the illumination of the test object in foot-candles. For the convenience of the operator in working on either side of the instrument, duplicate scales and pointers are provided on each side of the lamphouse. In order to eliminate slack or backlash in the gears and to provide an adjustment for wear, two gears are used at the end of each shaft, the purpose being that one shall be advanced just enough from the position of the other that all slack or looseness in the system shall be eliminated. To provide for this adjustment a short arc-shaped slot is cut near the edge of each outer gear through which a set-screw passes to the inner gear. In making the adjustment, the set-screw is loosened, the position of the outer gear advanced and the set-screw tightened. When positioned so that all looseness or slack in the transmission is eliminated, the setting of the vanes in any desired position can be made with great accuracy and precision.

For a later model, a simpler means was devised which is located within the housing and also is entirely free from slack or backlash. This is in the form of a thin plate with a central longitudinal slot through which pass the axle pins at the end of the vanes. Above and below it in alternate sequence are short vertical slots which engage pins suitably positioned at the ends of the vanes. The indicator which passes over the scale is attached to the end of the axle of one of the vanes. When this indicator is moved along the scale, the contiguous vanes rotate in opposite directions as is desired.

As mounted on the lamp arm of the perimeter, the attachment remains in fixed relationship to the perimeter arc at all positions of rotation of the arc. However, unlike the original device, it illuminates uniformly an arc of only 70 deg., the breadth of the tangent screen; therefore, the housing is provided with a swivel mount so that it can be turned on its vertical axis in order to illuminate the test object through the required 180 deg. of arc.

With this attachment the light and color minima can be quickly and easily determined at any point in the visual field. Also the field and the blind spot can be mapped for light and

color at different levels of illumination. This latter feature greatly increases the number of points in the field that can be tested by the method of limits, giving a much greater range and versatility in this respect than are given by variation in the size of the stimulus and its relation to the background. The attachment is inexpensive to make and easy and convenient to operate.

Some Preliminary Results

As already indicated, the attachment described can be used either for the determination of the light and color minima or as an auxiliary equipment to give different levels of illumination in the use of the method of limits for the study of the visual field. For convenience the latter will be discussed first.

1. As auxiliary equipment in the use of the method of limits the following are some of the points that may be noted:

(*a*) The form and color fields may be mapped at different levels of illumination. This is of value in two ways. In the first place it provides a means of testing the functional condition of the retina at as many points from center to periphery as may be desired. This is possible because the controls provide for changing the illumination in continuous series; thus intensities of illumination can be found which will cause the limits of sensitivity to change by as large or as small amounts as may be wanted. In the use of the method of limits in this way evidence as to pathologic disturbance, for example, may be found roughly in an inspection of the relationship between the limits obtained for the intensities used, or more accurately by a comparison of the results with predetermined norms or critical values for the limits for whatever intensities are selected for the purpose. In this connection it may be noted, however, that the method of limits does not provide as sensitive a means for testing the functional condition of the retina as the light and color minima for, as already shown, in some parts of the retina it requires very great losses of sensitivity to produce a change of 1 deg. in the location of

the limit. In Fig. 2, is given a field map showing the limits for blue (Hering pigment paper) obtained at different intensities of illumination. The range of change of intensity is greater than that provided with the attachment described, but the results serve to show the possibilities of testing the functional condition of the retina at different points by varying the intensity of the illumination.

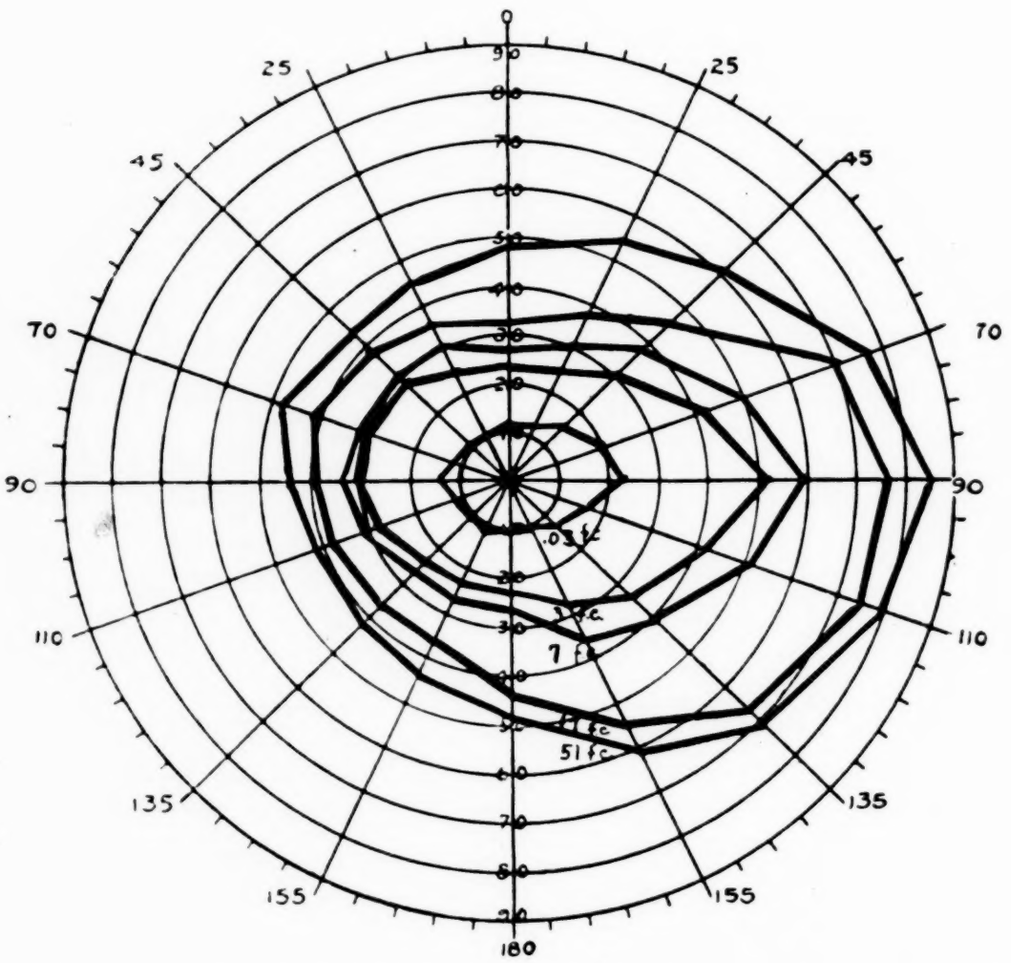


FIG. 2. Testing the functional condition of the retina at different points from center to periphery by the method of limits. Field maps showing the limits for a 1-deg. blue stimulus under 51, 17, 7, 3 and .03 ft-c of illumination. The background and preexposure were chosen in each case of the same brightness as the stimulus as seen at the limit of sensitivity.

In the second place, a knowledge of the breadth of field under different intensities of illumination has considerable value in relation to vocational fitness. For example, a test of the breadth of the field at one intensity alone is a very

inadequate check of the fitness of an aviator even for day flying and has very little value as a test of his special fitness for flying at night or under low illumination.

(*b*) The functional condition of the retina may be tested at different points in any given meridian, in any segment of that meridian, or in any part of the retina. As contrasted with the minima, the method in this case is to set the illumination at a given value and find the point in the field which is the limit of sensitivity for the stimulus used, instead of selecting a point in the field and finding the least amount of light that will render the stimulus visible at that point. Since the illumination can be changed in continuous series, a means is thus provided for determining as many limits in a given meridian or in a given sequence as may be wanted. In using the method of limits in this way, the results could be expressed in the form of a curve in which degree of eccentricity in the field would be plotted against intensity of illumination. In this case diagnosis, for example, would be made either by an inspection of the shape of the curve or by a comparison with previously determined norms or critical values. Again attention should be called to the fact that the method of limits is, for the reasons given in (*a*), not nearly so sensitive nor, without an enormous amount of work, so comprehensive as the method of minima for detecting changes in the functional condition of the retina, particularly in those portions of the retina where a test method having a high degree of delicacy is needed, namely where the gradient of sensitivity is steep such, for example, as in the extreme periphery of the field, around the blind spot and ordinarily around scotomas.

(*c*) The intensity control provides an additional means of adding sensitivity to the use of the method of limits for detecting disturbances which may cause a general contraction of the field, sector or regional cuts in the field, enlargement of the blind spot, or scotomas. Even with the addition of this means, however, the method of limits has much less sensitivity than the method of minima.

2. As providing for the direct determination of the light and color minima at any point in the central and peripheral

fields, the attachment affords for the first time in an instrument suitable for work in the clinic the most comprehensive and sensitive means known to workers in physiological optics for testing the functional condition of the retina; that is, the test can be made not only at every point in the field but also within the limits of delicacy of a just noticeable change in sensation.

In this use of the attachment the determinations may be made at isolated points or in any sequence of points that best serves the purpose of the examination. For example, they may be made at successive points from center to periphery in any meridian; or they may be limited to any part or segment of the meridian or to any sequence of points that will test the part or area of the retina that needs to be examined—such as the area immediately surrounding the blind spot, sector or regional cuts in the field, the region occupied by a scotoma, or zones or regions in which it is known that pathological disturbances are apt to occur. In all these cases the examination can be made with as much or as little detail as may be desired and with a delicacy that can be equalled by no other procedure. In all local disturbances the extent and profundity of the disturbance may be determined by a comparison of the results at the different points in the sequence.

As presenting some preliminary results obtained with the attachment in the determination of light and color minima, Figs. 3–6 have been prepared.[1] The curves shown in these figures afford some idea of the sensitivity gradient for the light and color sense from center to periphery, also of the deviations from this gradient which may occur as the result of pathological disturbance. The curves were determined in the 30-deg. upper temporal meridian of the visual field. This meridian was selected because of its importance in testing for certain pathologic disturbances.

The determinations were made in a darkened room. During the period of preparation the intensity of light on the perimeter arm was set at a low value (approximately .001

[1] The data used in the preparation of these figures were obtained under our direction by E. F. Lewis and D. A. Franklin.

ft-c) and the observer allowed to adapt to this illumination. This procedure provided for the standardization of the sensitivity of the retina in sufficient amount for the purpose and at the same time avoided the inconvenience of at any time

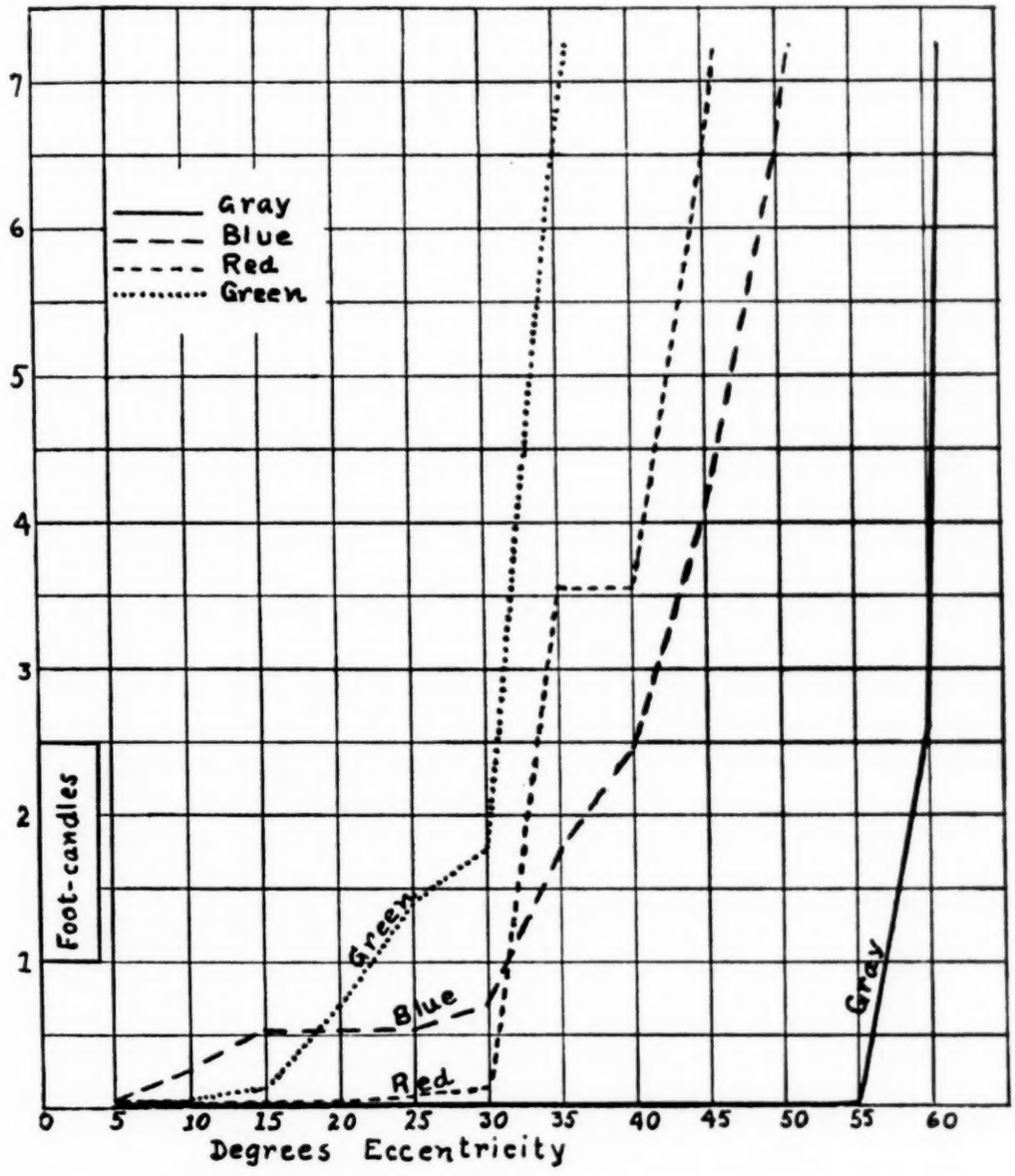


FIG. 3. The minima for red, green, blue and gray from center to periphery of the field of vision for one representative nonpathologic observer. Degree of eccentricity plotted against ft-c of illumination.

working in complete darkness. In making each determination the stimulus was moved to the appropriate position on the perimeter arc, the observer took his fixation and the light was quickly increased until an approximate value of the minimum

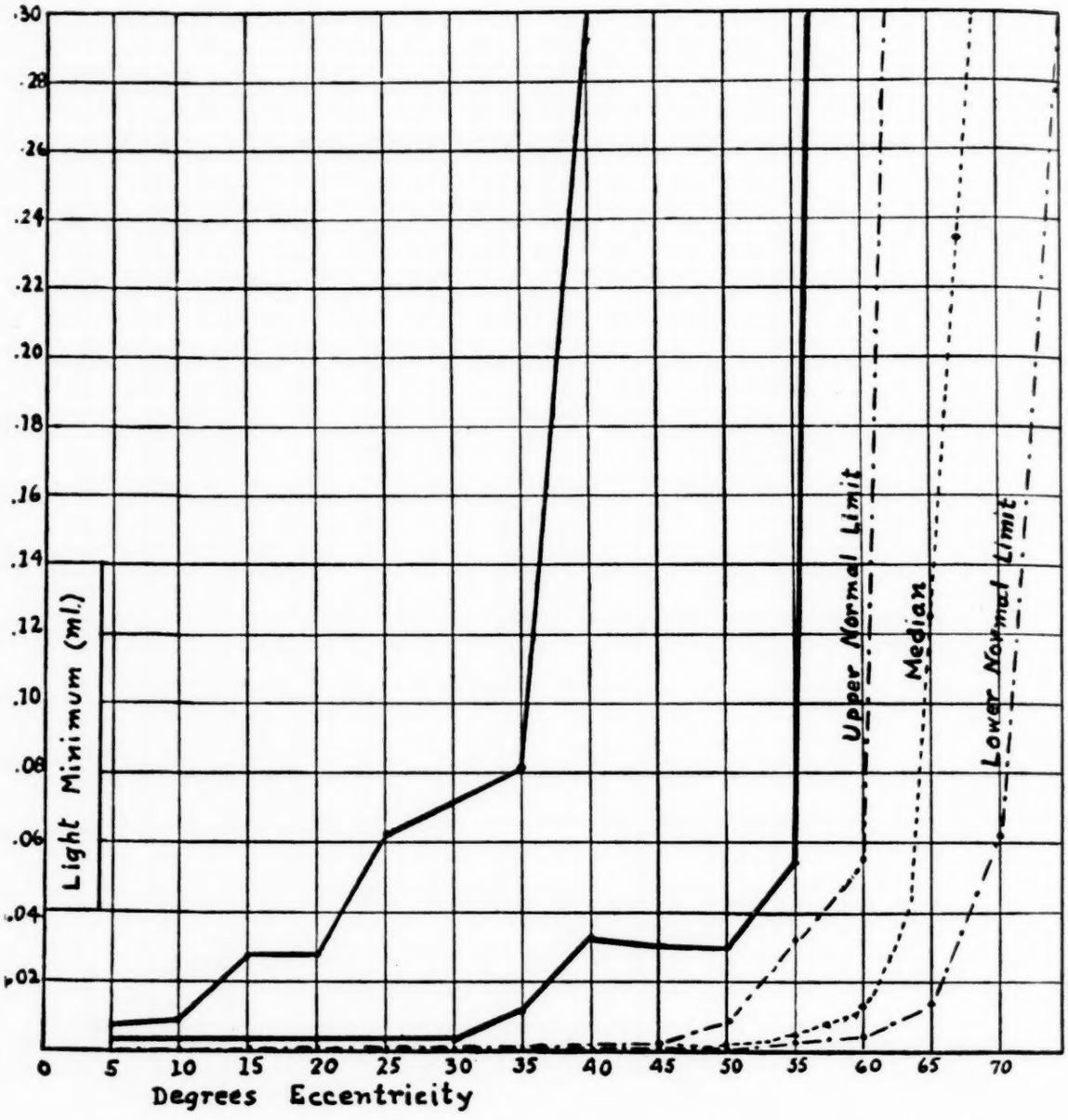


FIG. 4. The light minimum from center to periphery of the field of vision. Curves for the limiting and median values of a group of nonpathologic observers (broken lines) and for two pathologic cases (solid lines) in the 30 deg. upper temporal meridian. This meridian was selected because of its significance for diagnosis. Stimulus 1 deg. gray (reflection factor 8.5 percent) on black (reflection factor 4 percent). Degree of eccentricity plotted against millilamberts. Curves of this type are of particular value in a systematic study of the characteristics of eye diseases. When determined in a sufficient number of meridians, they show what parts of the retina should be tested.

was obtained. The value was then more accurately determined by using the method of preëxposure; that is, the stimulus was covered with the preëxposure card, the intensity of

the light quickly readjusted and the stimulus exposed at the new intensity for 1 sec.—the procedure being repeated until the exact amount of light was determined at which the stimu-

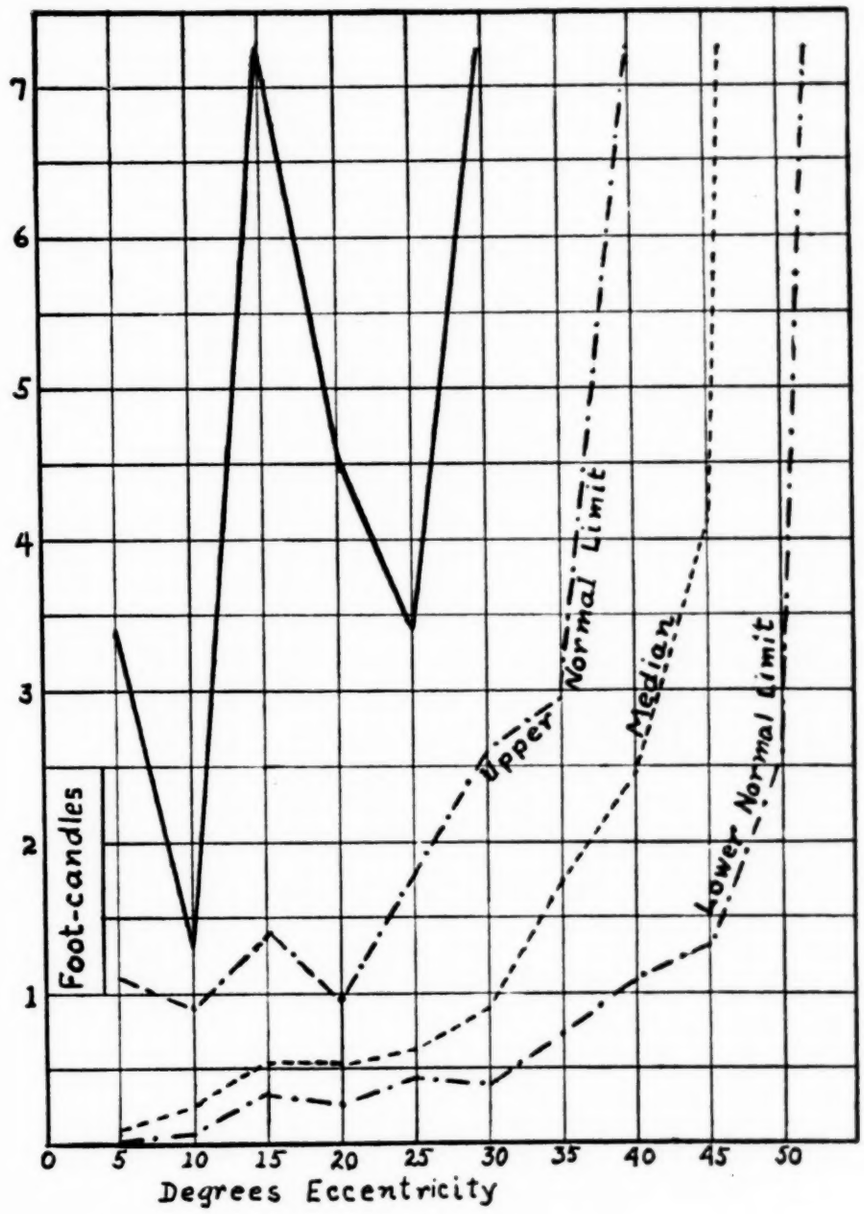


FIG. 5. The minimum for blue from center to periphery of the field of vision. Curves for the limiting and median values of a group of nonpathologic observers (broken lines) and for an early case of glaucoma (solid line) in the 30-deg. upper temporal meridian. Stimulus 1 deg. blue on gray (reflection factor 9.6 percent). Degree of eccentricity plotted against foot-candles of illumination.

lus was just seen. A practiced examiner operating the intensity control with one hand and manipulating the preëx-

posure card with the other can make these determinations very easily and quickly—much more easily and quickly than might be supposed.

The brightness of the preëxposure and surrounding field

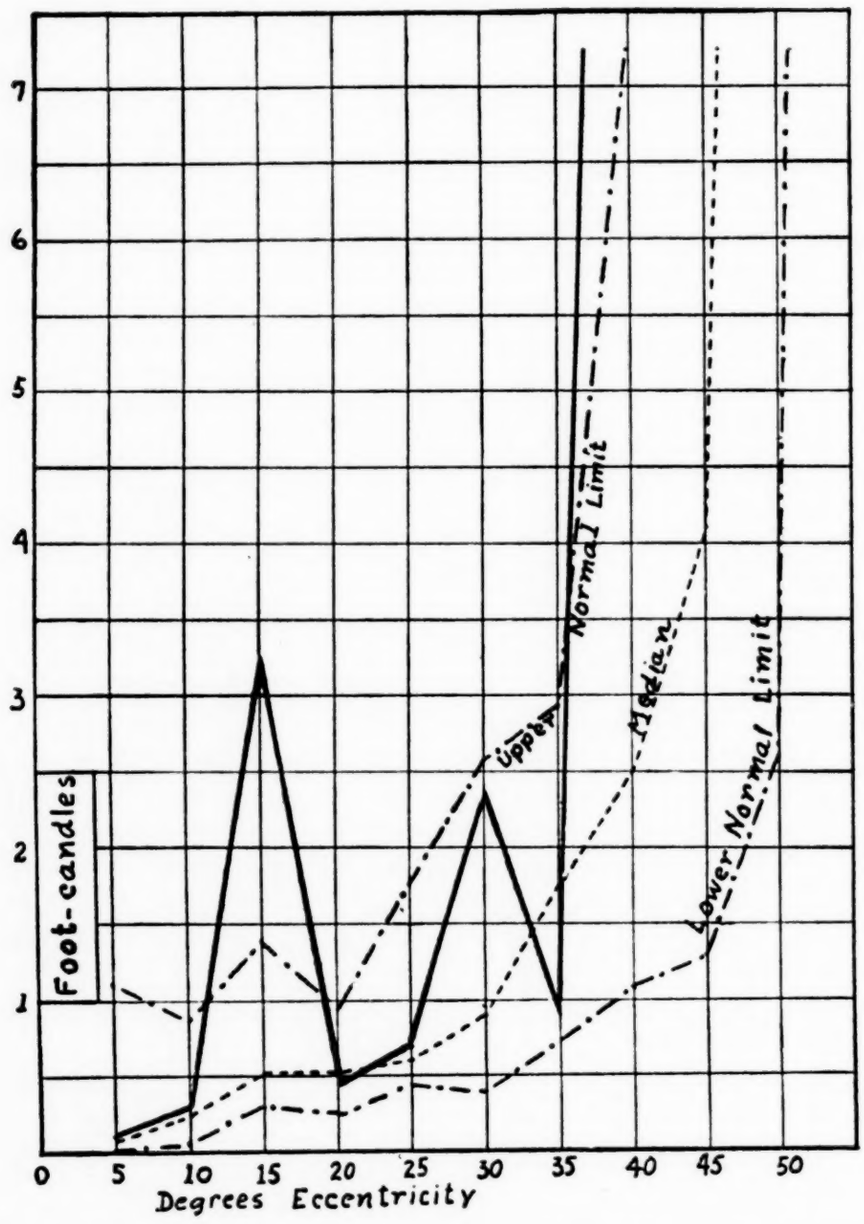


FIG. 6. The minimum for blue from center to periphery of the field of vision. Curves for the limiting and median values of a group of nonpathologic observers (broken lines) and for a case which has not yet been diagnosed as pathologic (solid line) in the 30-deg. upper temporal meridian. Stimulus 1-deg. blue on gray (reflection factor 9.6 percent). Degree of eccentricity plotted against ft-c of illumination. This figure has been prepared to show the irregularities that may be found in the minima in some cases in which a diagnosis of pathologic disturbance has not yet been made.

were the same as we have recommended and used in perimetry; that is, for the colored stimulus it was the same as that of the colors and for the colorless stimulus it was black. The reasons for making this choice of preëxposure and surrounding field have been fully discusssed in previous papers (**5**). Both the colored and the colorless stimuli subtended a visual angle of 1 deg. This is regarded by us as in general the best size to use. Where a still higher degree of sensitivity is desired, a smaller stimulus may be employed.

For the colorless stimulus a gray (reflection factor 8.5 percent) was chosen. For this choice the following reasons may be offered: (*a*) If a white stimulus were used the determinations would all be made at much lower intensities of illumination. Work at these lower intensities would be inconvenient, particularly for clinic work, also would cause the determinations for the light minimum to be made at a state of adaptation unduly different from that for the color minima. With the use of the gray stimulus chosen the determinations of the light and color minima are made at more nearly the same levels of adaptation—as nearly the same perhaps as it is possible to provide considering the way in which the determinations are made. (*b*) The gray stimulus requires a greater change in the amount of light to render it just visible from point to point in the visual field than a white stimulus. Therefore, the use of the gray stimulus gives the method of minima greater delicacy for picking up differences in sensitivity at different points in the field as measured by the amounts of light required to make the stimulus visible at these points. In other words, the use of the gray stimulus amplifies or magnifies the measuring scale. (*c*) The use of the gray stimulus does not require such fine changes in the setting of the intensity control. The use of a white stimulus, for example, would require such small changes in the setting of the control as to introduce undue and unnecessary difficulty both in the construction of the attachment and in its operation. A control which would produce such fine changes as would be required for a white stimulus and at the same time cover a range from 0 to 7 ft-c would be intricate in construction and

infeasible for clinic work. The use of the gray stimulus, on the other hand, renders the simple device we have described amply adequate on all these points—range of intensity covered, delicacy or sensitivity in making the determination, and ease of operation.

In Fig. 3 are given curves showing the minima for red, green and blue (Heidelberg pigments) and the gray stimulus previously mentioned. These results are for a single non-pathologic observer chosen as being as nearly representative as any we have ever examined. The minima are expressed in terms of ft-c falling on the pigment stimuli used. These results are given, it should be remembered, as illustrating those that may be obtained with the attachment to the Ferree-Rand Perimeter which we have described, and are intended only to be regarded as representative of conditions which are feasible for practical work. In previous papers have been given results obtained by us showing the sensitivity gradient from center to periphery in two meridians (**2**) and obtained under our direction in eight meridians (**6**) for red, green, blue and yellow. In both cases spectrum colors, filtered to give still greater purity, were used and the minima measured in energy units. Such results may be regarded as representative of the best type of scientific conditions of work and as having absolute value.

In Fig. 4 are given curves of the limiting and median values of the light minima from center to periphery for a small group of nonpathologic observers and, by way of comparison, the results for two cases of glaucoma.[2] These determinations

[2] Glaucoma has been selected for representation here because it is one of the most common causes of blindness and one that is attended by wide and varied disturbances in the light and color sensitivity of the retina, the mapping of which is one of the most important aids to the diagnosis of the disease and to the detection of its advance or recession (**7**). The disease is in general characterized by increase of intraocular tension or pressure which progressively destroys the functions of the retina. The disturbance usually begins in the region of the blind spot and in the far periphery on the temporal side of the retina. From these locations the disturbance spreads irregularly, presenting complicated patterns that make the disease particularly interesting from the standpoint of the study of the visual field. Fortunately the disturbance reaches the center of the field later, often at an advanced stage in the development of the disease.

Theories differ as to the cause of glaucoma. The increase of intraocular tension

were made in terms of foot-candles falling on the gray stimulus previously noted and the results converted into millilamberts for the sake of a more independent and more nearly absolute specification of the stimulus.

In Fig. 5 are given curves of the limiting and median values of the minima for blue for a small group of nonpathologic observers and, by way of comparison, the results for a case referred to us as an early stage of glaucoma. This case showed a contraction of the field, also a Bjerrum scotoma [3] for a ½ deg. blue stimulus but not for 1 deg.—the size of stimulus used in making the determinations for the minima. The marked deviations and irregularities in the curve for the minima will be noted, particularly the very high values of the minima in the region of the Bjerrum scotoma.

Figure 6 has been prepared to show the irregularities that may be found in the minima in some cases in which a diagnosis of pathologic disturbance has not yet been made. Whether these irregularities are due to the presence of more than usually large or dense blood vessels or to some other nonpathologic cause, or whether they should be regarded as indicating a very early stage of pathologic disturbance, we are not at this time prepared to say. The more striking irregularity in the case shown occurs, it will be noted, in the region of the blood-vessels emerging from the blind spot. This, it will be remembered, is also the region of the Bjerrum scotoma.

In the introduction we have given what we believe to be a fair evaluation in general of the possibilities of the use of the method of limits and of the light and color minima in field

---

seems to be due to a disturbance in the balance between the amount of fluid entering the eye and drained from it. Some are inclined to think that glaucoma may be present in cases where there is no detectable increase of intraocular tension, or at least no history of increased tension; others, that such cases although they may show disturbances in the visual field similar to those found where there is increased tension should not be classed as glaucoma. In any event it would be very difficult to explain the variety and complication of pattern of the areas of disturbance found in glaucoma and their differences from person to person and from time to time as an effect of pressure.

[3] The Bjerrum scotoma is an arc-like extension of the upper or lower border of the normal blind spot, the directional tendency of which is to encircle the paracentral field of vision. It was claimed by Bjerrum to be one of the earliest signs of glaucoma.

study. There remains to give a brief discussion of the value of the latter method in the present stage of development of clinic field study. While the use of the light and color minima opens a promising field for future development and should be of great service to the forward-looking clinic worker, its value at present doubtless is in supplementing the method of limits in cases where a more detailed study is wanted. The procedure used by many in examining a new case would probably be first to block out the situation roughly with the method of limits and then to use the light and color minima to complete the study where more detailed information is wanted. The most important service will probably be in the great sensitivity which the method of minima provides for detecting small changes in a disturbance, advance or recession, at the more critical and significant points in the field. Such uses of the minima do not require a prior establishment of norms and critical values and are, therefore, immediately available. In general it may be said too that in all phases of field study by the use of the minima, there is less dependence on a comparison with norms and critical values than there is in the method of limits. The showing made on the minima by pathologic disturbances in the peripheral retina is so great as to leave little doubt of their presence without a meticulous comparison with norms or critical values. In this respect there is a similarity to the certainty that is afforded by a sector cut in the field in the use of the method of limits. From all these considerations we believe that any competent and serious worker in clinic field study will readily see the advantage of having as part of his equipment the attachment for determining the light and color minima and for making perimetric and tangent screen studies at different levels of illumination.

The value and convenience of the attachment for certain types of research work and for laboratory and teaching work are equally obvious. The perimeter and its attachment makes of the study of light and color sensitivity a new subject and one that is readily available to every laboratory. It makes easily possible a wide range of studies in the peripheral field of vision not only of light and color sensitivity, their distribution

and gradients, but also of visual acuity and the effect of such factors as adaptation, intensity of light, etc. on these functions. At present, we scarcely need point out the peripheral field of vision is jungleland waiting to be explored and brought into scientific relationship with central vision.

With both the method of limits and the method of minima made more readily available for the purpose for which each is best suited and with the perimeter and its tangent screen standardized for the control of variable factors and provided with every known means for adding accuracy and delicacy in the use of the method of limits, together with the central vision scotometer (**8**) and the spectro-radiometric light and color sense tester and rotary campimeter (**9**) described in previous papers, we feel that a very great deal has been done, perhaps about all that can be done at this time, to render the study of the visual field complete, convenient and accurate.

Summary

The following methods have been used to measure light and color sensitivity: (1) The light and color minima and light and color differences. These methods are customarily used to measure light and color sensitivity in central vision. They have not been used for practical work in peripheral vision, however, because of the lack of a suitable and feasible means of presenting at all points in the field a stimulus whose intensity can be varied. (2) The limits of sensitivity (the perimetric method). In this method, the one customarily used to study sensitivity in peripheral vision, the points farthest from the center of the field are determined at which a given stimulus is visible. In the minima or threshold method, on the other hand, the intensity is determined at which a given stimulus is just visible at a given point in the field. The latter method is the older and without doubt the logical procedure; the former may be regarded as an indirect and substitute procedure, in many respects a poor one. (3) The light minimum for the discrimination of detail. The usefulness of this method in testing fitness for various kinds of work and types of ocular performance is obvious. So far

as we know, it has not yet been used to a considerable extent in clinic work, but we believe it should have a place in every comprehensive program for the study of the characteristics of eye diseases.

The shortcomings of the method of limits are discussed. Perhaps the most serious fault of the method is the examiner's inability to make the test where he wishes. Examples of the need of full freedom in this respect in the medical field are the early stages of detachment of the retina and during the course of reattachment; all disturbances in the central, paracentral and mid-peripheral portions of the retina; early stages of the disturbance in all parts of the retina and all cases of advance or recession of the disturbance. The complete adequacy of the method of minima in all such cases and the serious inadequacy of the method of limits should be clear to all. Examples of the need for freedom in other fields vary with the problem under study and are too numerous to mention here.

In recognition of the usefulness and need of the method of minima as well as the method of limits, an attachment to the Ferree-Rand Perimeter is described by means of which both the light and color minima can be easily and conveniently determined at any point in the field and the method of limits extended to include the use of stimuli of different intensities in a continuously graded series. Some preliminary determinations of minima made with this attachment are given to show sensitivity gradients for light and color from center to periphery and variations in these gradients which may be due to pathologic disturbances and other causes. A discussion is also given of the uses of the attachment as supplementary equipment for the study of the visual field by the method of limits.

The attachment is convenient and easy to use and can be readily substituted for the lamphouse of the perimeter. It has particular value (1) in all cases (*a*) where a more detailed study is needed than can be made with the method of limits; (*b*) where a means more sensitive than is provided by the method of limits is needed for detecting differences in sensitivity or changes in sensitivity from time to time—*e.g.*,

pathologic disturbances or advances or recessions in these disturbances; (*c*) where quantitative knowledge is wanted of the relationship of the sensitivity in the central retina to the sensitivity at any point in the peripheral retina or between different points in the peripheral retina, and the effect of adaptation on these sensitivities and relationships. This opens up a new field for both scientific and practical studies; (*d*) where knowledge is wanted of the limits of the field of vision under different intensities of illumination and the effect of adaptation on these limits. One example of this need is in testing the peripheral vision of night flyers; and (*e*) where an extension of the versatility of the perimetric method is wanted beyond what is given by variation in the size of the stimulus or in its color and brightness relation to the background. And (2) in a systematic study of the characteristics of eye diseases as the most thorough means of showing what parts of the retina are apt to be affected.

(Manuscript received July 20, 1939)

References

1. Ferree, C. E., and Rand, G., Factors which influence the color sensitivity of the peripheral retina, *Trans. Amer. Ophthal. Soc.*, 1920, **18**, 171–197; size of stimulus and brightness of preëxposure and of surrounding field in relation to the extent and shape of the color fields, *J. gen. Psychol.*, 1930, **4**, 15–35; two important factors in the size and shape of the form field and some of their relations to practical perimetry, *Ibid.*, 1932, **6**, 414–428.
2. Ferree, C. E., and Rand, G., Chromatic thresholds of sensation from center to periphery of the retina and their bearing on color theory. Part I, *Psychol. Rev.*, 1919, **26**, 16–41.
3. Ferree, C. E., and Rand, G., A new laboratory and clinic perimeter, *J. exp. Psychol.*, 1922, **5**, 46–67.
4. Allman, T. L., and Jenkins, P. H., A test for night visual efficiency, *Flight Surgeon Topics*, 1938, **2**, 40–46.

   Hargreaves, J. M., A test for nocturnal visual efficiency, *Ibid.*, 1938, **2**, 74–75.

   Platt, R. J., and Griffis, L. E., Report on test for night visual efficiency, *Ibid.*, 1938, **2**, 90–93.
5. Rand, G., The factors which influence the sensitivity of the retina to color: a quantitative study and methods of standardizing, *Psychol. Monogr.*, 1913, **15**, 166 pp. + xi.

   Ferree, C. E., and Rand, G., The limits of color sensitivity: effect of brightness of preëxposure and surrounding field, *Psychol. Rev.*, 1920, **27**, 377–398; size of stimulus and brightness of preëxposure and surrounding field in relation to the extent and shape of the color fields, *J. gen. Psychol.*, 1930, **4**, 15–35.

FERREE, C. E., RAND, G., AND MONROE, M. M., Critical values for the limits of the color fields in the eight principal meridional quadrants taken separately, *Amer. J. Ophthal.*, 1933, **16**, 577–589.

6. WENTWORTH, H. A., A quantitative study of achromatic and chromatic sensitivity from center to periphery of the visual field, *Psychol. Monogr.*, 1930, **40**, 189 pp.
7. FERREE, C. E., RAND, G., AND SLOAN, L. L., Sensitive methods for the detection of the Bjerrum and other scotomas, *Arch. Ophthal.*, N. Y., 1931, **5**, 224–260; Roenne's nasal step as studied with stimuli of different visibilities, *Ibid.*, 1931, **6**, 877–900; selected cases showing the advantages of a combined tangent screen and perimeter, *Ibid.*, 1933, **10**, 166–184.
8. FERREE, C. E., AND RAND, G., A central vision scotometer, *Arch. Ophthal.*, N. Y., 1933, **9**, 608–617.
9. FERREE, C. E., AND RAND, G., A spectroscopic apparatus for the investigation of the color sensitivity of the retina, central and peripheral, *J. exp. Psychol.*, 1916, **1**, 247–283; radiometric apparatus for use in psychological and physiological optics, *Psychol. Monogr.*, 1917, **24**, 65 pp. + xii.

IMPLICIT OR PARTIAL REVERSION-ERRORS: A TECHNIQUE OF MEASUREMENT AND ITS RELATION TO OTHER MEASURES OF TRANSFER [1]

BY ELSA M. SIIPOLA

Smith College

Introduction

This investigation concerns primarily certain methodological problems involved in the direct study of phenomena of 'transfer proper,' the actual carry-over of acts or items from an earlier performance to a later one. Such phenomena are to be distinguished from 'transfer-effects,' the effects of transfer as shown in speed of performance or rate of learning.[2] The cases to be dealt with here are those in which acts transferred from the earlier performance constitute errors in the later performance. The type of transfer involved here, transfer of a previously learned act to a later situation in which it is inappropriate, will be designated *reversion*.[3] This term will be used to refer specifically to phenomena of transfer proper. The terms *negative transfer* and *interference* will be used, in the customary manner, to refer to transfer-effects.

Reversion, as defined, can be dealt with experimentally and can be measured directly only by reference to actual reversion-

[1] This is part of an investigation presented to the Faculty of the Graduate School of Yale University in partial fulfillment of the requirements for the degree of Doctor of Philosophy in psychology. The writer is indebted to Professor Donald G. Marquis under whose direction these studies were made.

[2] The distinction between transfer proper and transfer-effects has been utilized recently by Woodworth. Woodworth, R. S., *Experimental Psychology*, New York: Holt, 1937. Pp. 176 ff.

[3] The term *reversion*, denoting actual relapse or going-back to an earlier performance, seems the most acceptable English equivalent for the terms used commonly in the German literature, *Rückfall* and *Rückfälligkeit*. *Regression*, in both German and English, is specific to the terminology of psychoanalysis in which it refers to broader phenomena of personality in relation to concepts of more primitive stages in development.

errors. Measures of transfer in terms of trials or time are indirect measures at best since they deal only with possible effects of reversion-phenomena as reflected in rate of learning or speed of performance.

Direct measurement of reversion has been accomplished only crudely by counting the frequency of overt reversion-errors. By this procedure reversion has been restricted experimentally to include only complete, literal reproductions of previous acts or items. A primary difficulty has arisen from the fact that such complete reversion-errors occur very infrequently.[4] However, even in experiments in which few or no overt reversions have occurred, other evidence has commonly indicated that a strong reversion-tendency was operating without being reflected in the measures used. Investigators have commonly noted the occurrence of such phenomena as preliminary starts toward execution of the earlier performance, erasures and corrections of first responses, and verbal comments reporting a 'change of mind' during the course of a response. Further evidence for the presence of a tendency to revert is found often in tense postural rigidity and increased verbalization, both of which reflect the excessive expenditure of energy by means of which the subject usually succeeds in avoiding overt reversions.

The hope of dealing directly with reversion seems to rest, then, upon the possibility of providing opportunity for reversion-phenomena, which are ordinarily implicit or not reflected in the measures used, to appear in recordable form. It must be made possible to identify experimentally those phenomena, notably the partial reversion-errors, which are truly reversions but which do not reach the gross level of complete, literal reproductions of the earlier performance. In order that quantitative measurement of reversion be attained, there must be developed a means of treating gradations in

[4] Preliminary experiments with verbal materials showed a negligible number of complete reversion-errors. For other evidence on this point see: Schwarz, G., Über Rückfälligkeit bei Umgewöhnung, *Psychol. Forsch.*, 1927, **9**, 86–158; 1933, **18**, 143–190. Siipola, E. M., and Israel, H. E., Habit-interference as dependent upon stage of learning, *Amer. J. Psychol.*, 1933, **45**, p. 226.

degree of reversion-error in addition to mere frequency of occurrence.

An attempt to devise an apparatus and a technique adequate to the detection and measurement of partial reversion-errors is reported here. A method of measuring reversion directly, incorporating the technique devised, is proposed. The relation of the proposed direct measure to reaction-time measures of reversion is determined by application of the present technique to an experimental situation in which direct comparison of the different measures is possible.

Apparatus and General Method [5]

The following requirements, which seem essential to the detection and measurement of partial errors, were taken into account in the design of the apparatus and the choice of a technique. First, the responses utilized must be of a kind which can occur in varying degrees of accuracy and completeness—they cannot be simple reactions of all-or-none character. In the present apparatus extended manual movements of a lever through an open, circumscribed area to certain terminal positions were utilized to meet this requirement. Second, a complete and detailed record of each response must be obtainable in order to measure the degree of accuracy or completeness. By the system of recording devised here a complete tracing of the path of movement for each reaction was secured. Third, there must be available a number of separate, comparable responses. Assignment of a specific terminal position for each separate response provided a whole system of comparable movements identifiable according to the radial direction of the lever-movement. The present experimental situation involved learning to respond with a particular movement to each of a set of visual stimuli, in the manner of code-learning. Fourth, the experimental situation in general must favor the occurrence of reversion-errors. The following provisions contributed toward this end: choice of a simple, rapid movement, likely to become relatively involuntary; presence of the open area allowing freedom of movement; complete unawareness on the part of the subject that his movements were being recorded.

Experimental Situation.—The subject was seated in a separate compartment from which none of the recording mechanisms or connections were visible. Directly in front of him at table-height was the horizontal reaction-plate through which the movable lever projected upward (Fig. 1). Above the center of the plate and at eye-level was an exposure aperture through which the visual stimuli appeared.

The flat reaction-plate was made of thick hard rubber, cut out to provide an open semicircular reaction-area within which the lever could move freely. The straight (10 in) diameter toward the subject was broken at the exact center by a narrow channel which served as the starting position for each reaction. Along the curved periphery of the semicircle were 20 regularly spaced radial slots providing equidistant terminal positions for the different reactions.

---

[5] The writer is indebted to Professor Harold E. Israel of Smith College for assistance in the design and construction of the apparatus.

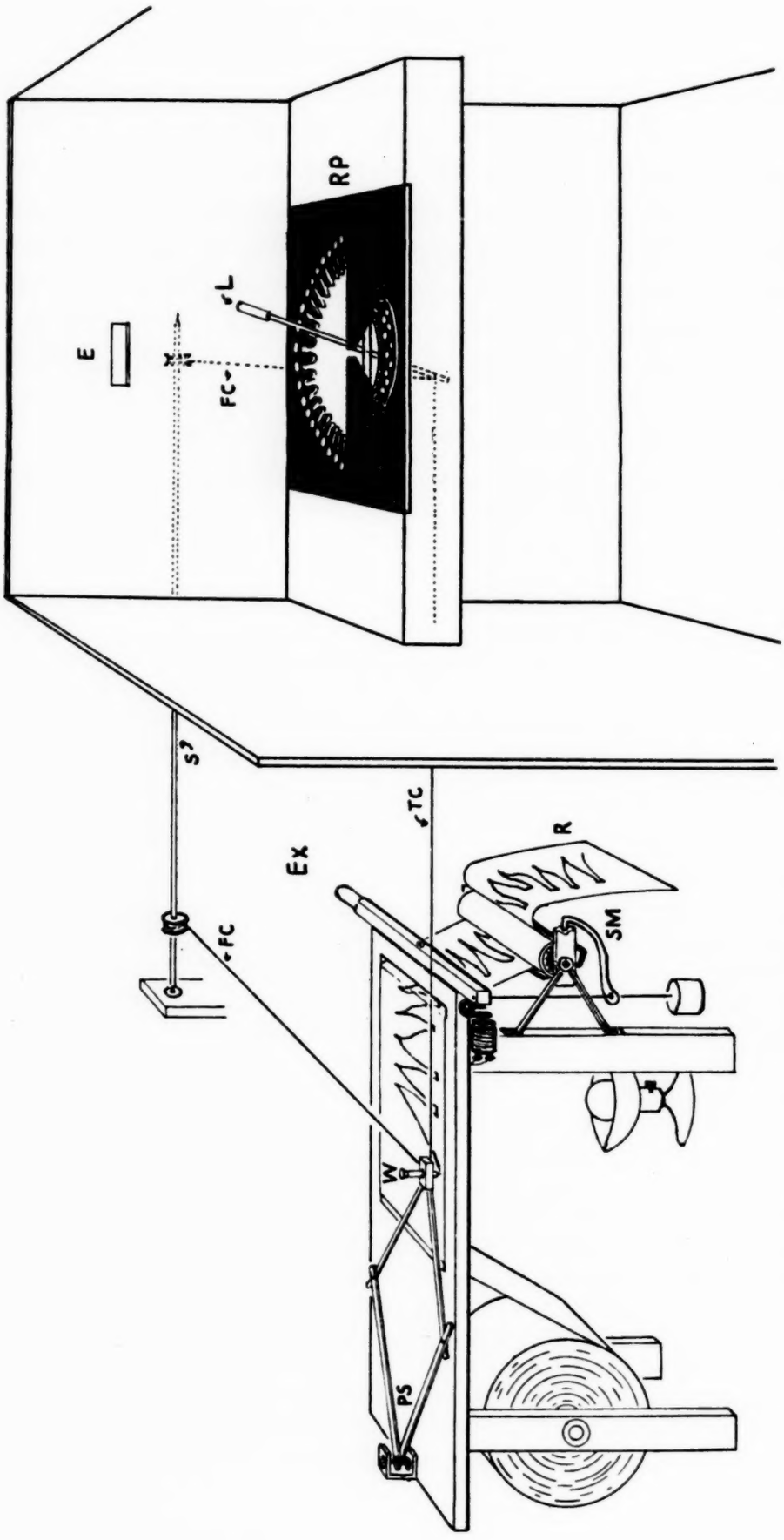


FIG. 1. Diagram of the apparatus. *RP*, reaction-plate; *L*, lever; *E*, exposure aperture; *TC*, transverse connecting cord; *FC*, forward cords; *S*, shaft; *W*, writing point; *PS*, pencil support; *SM*, paper-shift mechanism; *R*, record; *Ex*, experimenter's position.

The lever, a quarter-inch brass rod with rubber handle, extended vertically seven inches above the plate and had its fulcrum fixed directly below the starting position. The lever was pivoted through a slit in a heavy transverse axle, this suspension serving as a universal joint to allow free radial movement in all directions. At the extreme positions (in the radial reaction-slots) the lever was inclined 73° from the vertical. Within the reaction-slots the extent of movement was only 5°, allowing freedom throughout most (68°) of the course of each reaction-movement. Such an area of free movement was an essential feature of the present technique which involved the detection of variations in the paths of movement for the identification of partial reversion-errors.

Throughout, the task of the subject was that of learning to respond with a different reaction-movement to each of a set of visual stimuli, a set of such movements with their associated stimuli constituting what might be called a 'motor code.' For each reaction the sequence of events was as follows: At a ready signal the subject moved the ever from a resting position to the starting position, automatically sounding a buzzer upon arrival. After an interval of two seconds the stimulus was exposed by the opening of a shutter. As quickly as possible the subject reacted by moving the lever to the appropriate reaction-slot. Arrival at the end of the reaction-slot sounded another buzzer which was the signal for returning the lever as quickly as possible to the starting position. Each complete reaction included, then, movement from the starting position to the end of a reaction-slot and back to the starting position.

In order to standardize the procedure and timing during the learning of the motor code, a mechanical means for prompting was provided. The channel which included the starting position extended toward the subject and widened into an open prompting sector with contact-points arranged along its arc. Bringing the lever into contact with the point marked for a given stimulus revealed, by illumination from below, the designation of the appropriate reaction-slot. In order to discourage continued reliance upon prompting, a delay of five seconds before prompting would occur was mechanically provided.

Exposure and Timing Devices.—Exposure of the stimuli was controlled by a disc shutter behind the aperture in the screen facing the subject. Opening of the shutter by the experimenter automatically closed a circuit to start a Springfield Timer. Return of the lever to the starting position automatically closed the shutter and stopped the timer. Thus the reaction-time for a single reaction was always measured from the exposure of the stimulus to the return of the lever to the starting position. For convenience in handling the large number of stimuli required for an experimental setting, lists were prepared and attached to a Chicago type memory drum placed behind the exposure aperture and operated manually.

Method of Recording.—The system of recording was designed for the essential purpose of producing a single, two-dimensional copy of the original movements, a type of record easy to read and simple to interpret. A literal tracing of each reaction-movement was obtained by means of a system of direct mechanical connections between the extended lower end of the reaction-lever and a distant writing point. Through a parallelogram of connections, two dimensional components of the lever movement were transmitted to a single writing point which traced on a flat surface a copy of the reaction-movements. The connections were linen cords, one leading from the lever in the transverse direction directly to the writing point, and the other leading forward to a fixed pulley on a transverse shaft from which another cord led back to the writing

point (see Fig. 1). Counter-weights were applied to balance the system and to produce equal ease of movement in all directions.

For the detection of partial errors, it seemed highly desirable to provide a recording device which would be more sensitive to movement near the starting position than to movement near the terminal slots. Radial movement of a lever was chosen with a view toward securing conveniently such differential sensitivity. By reduction of the three-dimensional movement at the end of the lever to the two-dimensional movement of the writing point over the flat recording surface, a record was obtained in which increments of lever-movement from the starting position were represented by distances whose values followed the sine of the angle of inclination. The adjustment was such that one degree of inclination from the vertical starting position was represented at the writing point by a distance precisely three times that which represented one degree of inclination at the terminal slots (near the horizontal). By making the distance between lever and writing point relatively large (60 in) in comparison to the radius of the lever (5 in), distortion of the true sine-function was reduced to an insignificant factor (less than 0.5 percent at its maximum).

This design for the recording system placed severe demands upon the writing mechanism which had to be capable of extremely free and rapid movement of relatively great extent. In recording a single reaction the minimal excursion of the writing point was five inches, and a complete reaction (excursion and return) occurred commonly in less than one second. The device finally adopted was a mechanical pencil loaded with indelible lead and marking upon ordinary white roll-paper; the indelible lead moved smoothly over the surface with little friction and it left a tracing which could be made permanent by steaming. The pencil was held upright by a light wooden planigraph structure delicately pivoted to allow maximal freedom of movement.

In accordance with the essential design for obtaining a direct copy of the reaction-movement, a record which would remain stationary during the tracing was prescribed. It was necessary, however, to find some means of separating the outward excursion from the return movement and of separating successive reactions. The means adopted was that of producing a quick shift of the paper, timed automatically to occur at the moment when the lever completed an outward movement at the end of a reaction-slot. The shift was accomplished by the use of a typewriter carriage operated by the release of a weight when the lever made contact with the end of a slot. The record paper (9 in wide) was led across a ground-glass recording surface and on through the typewriter carriage. The glass bore a permanent 'map' corresponding in terms of the record to the entire area of the reaction plate, and it was illuminated from below to enable the experimenter to follow the position of the lever at all times.

The kind of record obtained can be described most conveniently by reference to Fig. 2 representing a sample of two complete reactions in succession. In the record, the starting position is at the top of the straight vertical lines which represent movements along the narrow channel of the plate. At the left, the first long line curving upward represents the first outward movement to one of the reaction slots. From the terminal point the short, straight, transverse line marks the shift of the record-paper. From the right end of the shift-line the tracing of the return movement descends to the starting point, now in a new place on the record and separated from its former place by the distance of the shift. The record of the next reaction to a different reaction-slot starts from this point.

The reading and analysis of the records was accomplished by placing the tracings

over the accurate 'map' on the ground-glass recording surface. This map had been constructed from actual tracings of the lever movements along a ruler placed on the plate in the direct path from starting position to each of the reaction-slots. Thus any slight constant distortion in the recording could not be reflected in measurement of the tracings. For matching each tracing to the points on the map, the vertical lines representing movement in the narrow channel and the position of the terminal point provided means of precise alignment. Measurement of variations in the paths of movement from the direct path could be made readily in terms of angular units. The reliability of the recording system could be judged by tracing numerous reaction-movements without shifting the record; repeatedly moving the lever along direct paths (along a ruler on the plate) gave single, superimposed records with no appreciable error.

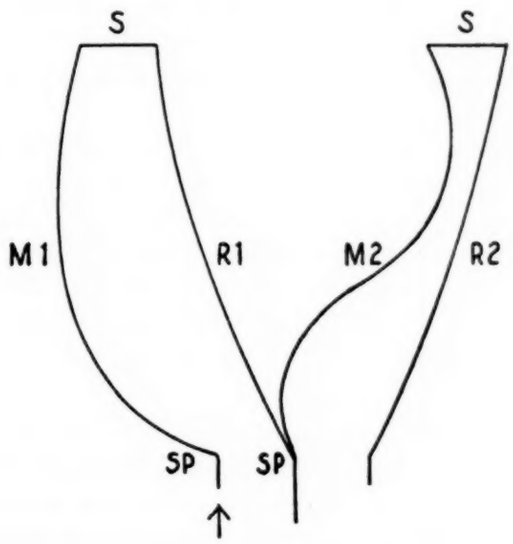


FIG. 2. Sample record representing two complete reactions (⅓ size). *SP*, starting position; *M*1, *M*2, outward reaction-movements; *R*1, *R*2, return movements; *S*, shift of record.

THE EXPERIMENT

The present experiment was designed to provide conditions favorable to the occurrence of habit-reversion so that the adequacy of the technique proposed for the direct measurement of reversion-errors could be tested. The primary purpose was the establishment of the validity of the assumptions basic to the new technique introduced, namely the proposals that partial errors are prominent phenomena of habit-reversion, and that by application of an appropriate technique reversion-errors can be reliably recorded and measured to yield a sensitive measure of reversion-tendencies. Provision was made for securing data of a type to which the standard indirect measures of reversion by reaction-time could also

be applied since the value of introducing a new technique of measurement must rest upon its possessing greater sensitivity or other qualities superior to those of measures already available.

Procedure.—The subject learned the simple place-habit of moving the lever to a certain reaction-slot, designated by a Roman numeral, when a certain Arabic number was presented as the stimulus. Ten stimulus-numbers (1–10) and ten of the reaction slots (every other one) were used; the slots were labelled (I–X) in regular ascending order beginning at the left.

In the first task, Task A, the subject was required to move the lever in each case to the slot designated by the Roman numeral corresponding directly to the stimulus-number. In each trial a single stimulus-number was exposed (*e.g.*, 4), and the task of the subject was simply that of moving the lever as quickly as possible to the numerically equivalent reaction-slot (*e.g.*, IV).

In the succeeding task, Task B, the conditions remained the same except for change of the responses assigned to two of the ten stimuli. The subject was directed: (*a*) when the stimulus 3 appeared, to move the lever now to VII rather than to III; (*b*) when the stimulus 8 appeared, to move the lever now to IV rather than to VIII; (*c*) for all of the other stimuli, to respond as before by moving the lever to the slot with numerically equivalent designation. It should be noted that slots IV and VII were now assigned each to two stimuli; although implied in the instructions, this point was not called specifically to the subject's attention.

The change of only two items within a ten-item code was chosen because this arrangement had been shown in preliminary experimentation to favor the production of reversion errors and because the use of only two critical items simplified greatly the task of analyzing the records. The choice of the particular items made the critical ones involved the following considerations. In order that the movement assigned in Task B should be easily distinguishable from that previously assigned in Task A, the two slots assigned to each critical stimulus were chosen to fall in different halves of the reaction-area, separated by 72° in each case. In order that reversion-errors should be represented by movements in both directions, the critical stimuli and slots were so chosen that reversion to the previous movement would involve movement to the right in the case of one stimulus and to the left in the case of the other.

Experimental Series.—For Task A the experimental series consisted of 75 reactions with the stimulus-numbers arranged in random order. The numbers 3 and 8 each appeared 15 times to insure a reasonable degree of training for these critical items. Stimuli 4 and 7 each appeared 10 times to provide data on the normal paths of movement to slots IV and VII, the slots assigned to the critical stimuli in Task B. The remaining stimuli were used to complete the series of 75 trials.

For Task B the experimental series consisted of 50 reactions again in random order. The critical numbers 3 and 8, now assigned new responses, appeared each 10 times to provide adequate opportunity for reversion. The numbers 4 and 7, which now shared assignments of slots IV and VII with the critical stimuli, appeared each 6 times. The remaining stimuli were used to complete the series of 50 trials.

In each trial the stimulus remained exposed until the subject closed the shutter automatically by the return of the lever to the starting position. The ready-signal for

the next trial followed immediately. An interval of three minutes separated the two tasks, and this interval was occupied completely by the giving of the instructions for Task B.

Subjects.—Fifteen subjects, all right-handed, served in the experiment. They were Smith College students selected at random from a large introductory class in psychology.

General Instructions.—At the beginning of the experiment the subject was given careful directions for the use of the apparatus, including the manner of holding and moving the lever and the sequence of movements to be made for each reaction. Preliminary practice trials were provided, ten for each reaction-movement to be used.

Certain features of the general instructions were important in connection with this experiment. The subject was told that the experiment concerned speed of reaction and that his time-score would be told him after each trial. The subject's interest was, thus, entirely centered upon arriving at the assigned slot with greatest possible speed. It was also of extreme importance that all subjects remained completely unaware of the fact that a record of the paths of movement was being made. (The subjects were all women and confirmed the common notion that the sex is completely lacking in curiosity concerning mechanical devices.) Double responses were effectively prevented by instructing the subject that once he had made contact in a reaction-slot (to sound the buzzer) his choice for that trial was completed and he was to return directly to the starting position. Partial errors and corrective movements were, therefore, limited to the phase of the reaction preceding the decisive contact. As a result of the instructions, then, the subject concentrated upon speed and was completely oblivious to any significance of the paths of movement followed through the free area. The fact that complete, overt errors in the choice of a reaction-slot were not called to the subject's attention reinforced the emphasis upon speed at the expense of accuracy.

Results

The Direct Measurement of Reversion in Terms of Errors.—The analysis of the results will be primarily concerned with the reactions to the two critical stimuli 3 and 8 for which the assigned movements were changed from III and VIII in Task A to VII and IV in Task B. Deviation of the critical reaction-movements in Task B in the direction of the paths previously learned was the aspect selected for measurement as a basis for detecting the presence and the strength of reversion-tendency. For each critical reaction the deviation of the recorded path was measured in terms of the angle formed by a line representing the direct path to the correct slot and a line from the starting position through the point of widest excursion toward the previously learned path.

In addition to measuring the deviations, it was necessary to establish, for the reaction movements (VII and IV) newly assigned to the critical stimuli in Task B, the limit of deviation

under normal conditions. Norms were obtained from the reaction-paths to slots VII and IV in Task A by measuring the deviations in the critical direction and calculating the 90th percentile score-values.[6] Both norms so calculated were identical in value, 19.8°. Reactions to the critical stimuli in Task B were, therefore, regarded as significantly influenced by reversion when they deviated in the reversion-direction beyond 19.8°, the limit established for normal reactions.

Before presentation of the quantitatitive treatment of reversion-errors, the nature of the reversion-movements themselves deserves brief mention. The experimental situation was such that the errors were major distortions in the path of movement, extending well along toward the reaction-slots, rather than minor tremors or hesitations near the starting position. The records indicate clearly that under the particular conditions the subject started immediately and moved so rapidly into the free area that any movement-tendency registered itself as an extended path. It is, therefore, highly unlikely that partial errors occurred without being readily detectable in the record.[7] Although there were minor variations, the paths containing partial reversion-errors in almost all cases (98 percent) followed a smooth path directed toward the reversion-slot and then shifted, either abruptly or by a swinging curve, toward the path into the correct slot. The complete absence of compromise-movements and the extreme infrequency of oscillating, conflict-movements is striking. The consistency in the manner in which the reversion-tendency manifested itself in this experiment may be a function of the particular conditions, or the consistency and the

[6] The type of norm desirable here should represent the limit beyond which errors seldom occur. For normal distributions the inter-percentile range having minimal error is that between the 6.9 and the 93.1 percentiles, as has been shown by Kelley, T. L., *Statistical method*, New York: Macmillan, 1924, p. 75. Since the present distributions were skewed, the 90th percentile score-value was adopted as a sufficiently close approximation.

[7] That the character of the movements and the errors was determined by the speed of reaction possible to the subject is amply supported by data showing that the character of normal movements is predictable from the degree of readiness or preparation possible to the subject.

form of manifestation may be more broadly characteristic of reversion-phenomena.

Figure 3 presents the distributions of the extents of devia-

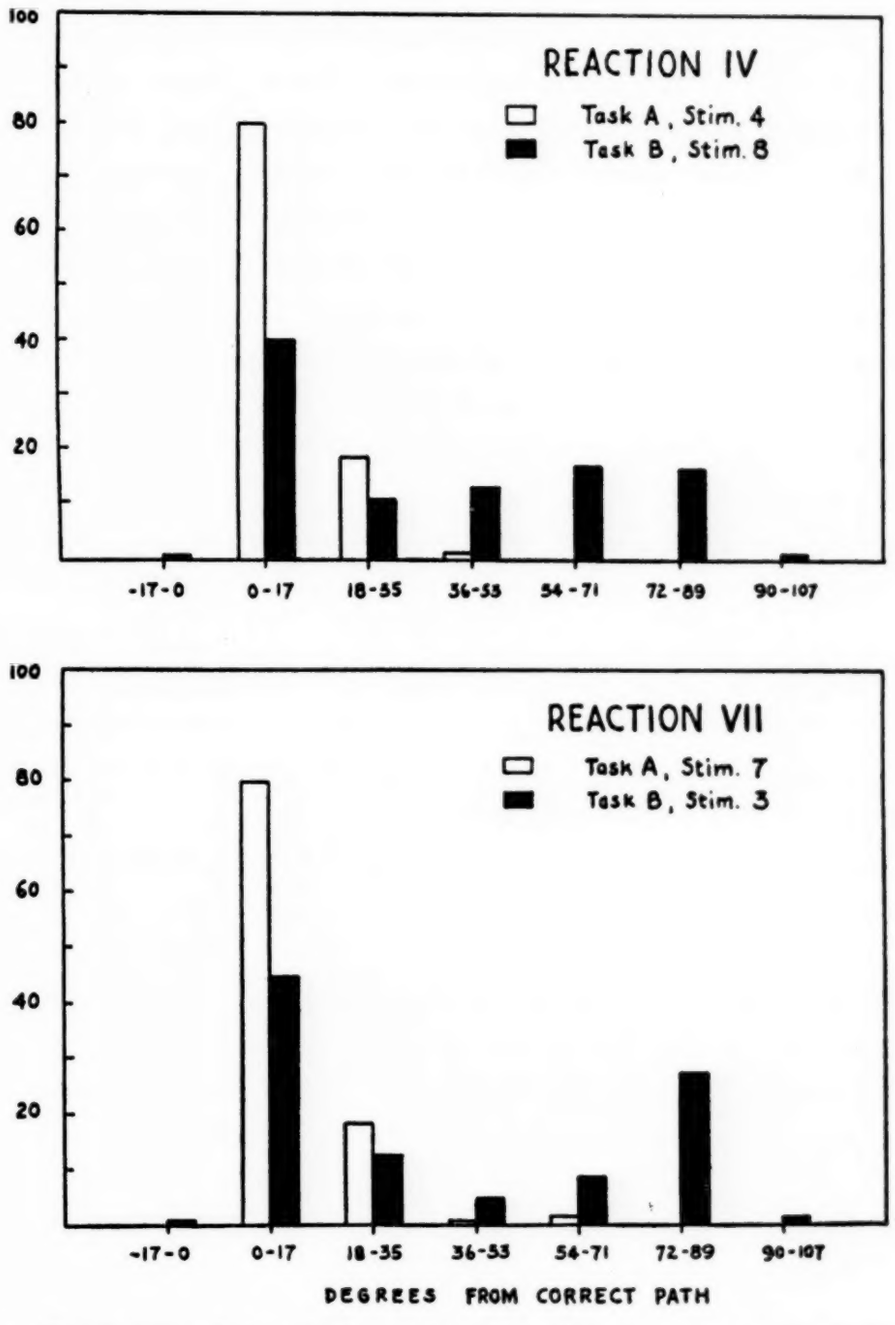


FIG. 3. Percentages of the reaction-paths which show various degrees of deviation (plotted in 18° intervals) toward the reversion-path located at 72°.

tion from the direct path measured in the critical direction. For both critical reactions, IV and VII, the normal paths of movement during the original learning of Task A deviated,

in a large majority of the cases, only within 18° of the direct path. The mean extent of deviation for these normal reactions of Task A was relatively small, 6.1° for reaction IV and 7.4° for reaction VII. For Task B, with the same reactions now assigned to new stimuli, the ranges and shapes of the distributions are clearly different. Here fewer than half of the reactions remained within 18° of the direct path, and the remaining cases were distributed among successively more distant intervals with a peak in the region of the reversion-path (72°) where the distributions end abruptly.[8] The mean extent of deviation for these critical reactions of Task B was relatively large, 32.9° for reaction IV and 32.8° for reaction VII.

The point of greatest interest in the distributions of Fig. 3 is the establishment of the fact that in a reversion-situation errors in the direction of the reversion-movement occur in all degrees with the frequency of the partial errors greater than that of the complete, overt reversions. This is even more true than is evident from the diagram since a considerable number of the cases which appear at the interval containing the reversion-path are actually cases of partial error corrected before the completion of the reaction.

The percentages of reactions significantly influenced by reversion, as indicated by comparison with the norms established for the same movements in Task A, are presented in Table 1. The treatment of the results is that of grouping together all of the reactions to each critical item of the code rather than that of combining averages for individual subjects, since the former procedure seemed clearly more pertinent. Figures in the total-column include all cases of reactions for which the path deviated in the critical direction beyond 19.8°, the limit established for the variability of normal movements. Under *complete errors* are included all reactions which terminated at the reversion-slot, either entered the slot precisely (sub-class, *precise*) or entered the immediately

[8] The difference between the distributions for the two reactions is due to the different proportions of complete and partial reversion-errors in the two cases. See Table 1.

adjacent slot (sub-class, *adjacent*). Under *partial errors* are included all reactions which terminated in the correct slot but deviated en route in the critical direction beyond the normal limit.

Comparison of the total percentages of significant deviation in the two tasks shows clearly the influence of reversion. In Task B half or more of the reaction movements were significantly distorted. The critical ratios for the differences between the two tasks are 9.2 for reaction IV and 8.4 for reaction VII. Analysis of the results for each of the individual subjects confirmed the presence of strong reversion-tendencies; for all but one of the fifteen subjects the proportion of significant reactions was in Task B far above the norms of Task A.

TABLE 1

PERCENTAGES OF REACTIONS SHOWING DEVIATIONS OF SIGNIFICANT EXTENT TOWARD REVERSION-PATHS

	Reaction	N	Totals of Complete and Partial Errors	Complete Errors		Partial Errors
				Precise	Adjacent	
Task A	IV	150	10.0±2.4	0.0	0.0	10.0
	VII	150	10.0±2.4	0.0	0.0	10.0
Task B	IV	150	53.7±4.1	6.7	5.3	41.7
	VII	150	50.0±4.1	18.0	3.3	28.7

The most striking point in the results presented by Table 1 is the large proportion of partial reversion-errors in contrast to the relatively small number of complete reversions. Approximately two-thirds (68 percent) of all of the significant errors were partial errors, whereas only about one-fourth (24 percent) were complete reversions of the precise type.[9]

The difference which appears in Task B between the two reactions in regard to the relative proportions of partial and complete reversion-errors has important bearing upon a relationship which appeared again commonly in later studies.

[9] The relative proportion of partial and complete errors is, naturally, a function of the particular experimental conditions. That the present conditions favored complete errors was shown by later similar experiments in which slightly changed conditions (shifting all ten items of the code) produced much higher proportions of partial errors, as high as 98 percent.

Reaction IV showed a much higher proportion of partial errors and a much lower proportion of complete errors. The proportion of the total number of reversion-errors (both partial and complete) in relation to the number of normal reactions remained, however, practically identical for the two critical reactions. This suggests the existence of a complementary relationship between partial and complete reversions. If such a relationship exists commonly, the measurement of strength of reversion-tendency in terms of only one of the types of error is obviously unsound. In the case of the present experiment, certainly, it would be concluded on the basis of complete errors only that a stronger reversion-tendency was present in the case of reaction VII, whereas an opposite conclusion would be reached on the basis of partial errors. Consideration of both types of error leads to the conclusion that the reversion-tendency was equally strong in the two cases.[10]

The results considered so far have been those for only the two critical items of the code, those involving changed assignments in Task B. It was apparent in the records that the effects of the reversion-tendencies in Task B extended to the reactions to two of the non-critical stimuli, 4 and 7. As in the case of all non-critical stimuli the reactions assigned to stimuli 4 and 7 remained the same throughout both tasks, but in the case of these two stimuli their assigned reactions were also assigned in Task B to the critical stimuli. Except in the cases of 4 and 7, the reaction-movements for the non-critical stimuli showed no appreciable disturbance in Task B.

The results for stimuli 4 and 7 in Task B show clearly distortions which are not mere inaccuracies or diverse disturbances; the movements tended distinctly to depart from the correct path into the direction toward certain specific reaction-

[10] Introspective comments suggest a reason for the difference in the proportions of partial and complete reversions for the two critical reactions. When the assignment for stimulus 8 was changed to IV, the subjects evolved a simple formula, "8 is now one-half of VIII, or IV," which operated quickly enough to prevent overt entrances into slot VIII but did not check in time preliminary starts toward this slot. The numerical relation between stimulus 3 and its changed assignment VII was not adaptable to such a simple formulation, and in this case the reversions tended to be complete.

slots. Reaction-movements for stimulus 4 deviated toward slot VIII, and those for stimulus 7 deviated toward slot III. These deviations represent movements toward slots whose assignment to these two non-critical stimuli *would have been* complementary to the actual assignment in Task B of slots IV and VII to the critical stimuli 8 and 3. Since the errors were as specifically directed as were those for the critical stimuli, the paths of reaction to stimuli 4 and 7 could be measured and analyzed in the manner already used for the critical stimuli. These results are presented in Table 2.

The results show that for these non-critical stimuli the percentages of significant errors are of about the same order of magnitude as those for the critical stimuli, differing reliably

TABLE 2

PERCENTAGES OF REACTIONS TO STIMULI 4 AND 7 SHOWING DEVIATIONS OF SIGNIFICANT EXTENT TOWARD PATHS VIII AND III

	Reaction	N	Totals of Complete and Partial Errors	Complete Errors		Partial Errors
				Precise	Adjacent	
Task B*	IV	90	40.6±5.2	5.6	3.3	31.7
	VII	90	48.9±5.3	8.9	6.7	33.3

* Results for Task A are those presented in Table 1.

from the results for normal reactions of Task A. Critical ratios for the differences between the two tasks are 5.3 and 6.8 for reactions IV and VII respectively. The differences between the results for the critical and the non-critical stimuli are statistically unreliable, indicating that the strength of the inter-task influence is of the same order for both types of stimuli.

Despite the similarity in quantitative aspects, the inter-task influences upon the reactions to the critical stimuli and those to these non-critical stimuli involved seemingly opposed mechanisms. In the case of the critical stimuli, the errors involved actually going back to the previously learned mode of reaction; for the two non-critical stimuli, the errors involved departure from, rather than return to, the previous mode of

reaction, even though the reactions previously learned remained the correct ones. The mechanism in the latter case seems to involve what might be described as a tendency for items, closely related to those in which changes are specifically required, to be influenced toward change in the direction of producing in the new habit a consistent system of reactions closely parallel to the system previously learned. The phenomena here raise serious difficulties for the explanation of reversion in terms of the separate, specific items of association theory or conditioning. The phenomena are essentially systemic in nature and are more readily assimilable to theoretical concepts of systemic trace-formation in learning. Practically, and from the point of view of methodology, the phenomena are extremely important since gross measures of transfer-effect are as much influenced by these indirect phenomena as by direct, literal reversions.

The Direct Measurement of Reversion in Terms of Errors as Related to Indirect Measurement by Reaction-Time.—The suggestion that a direct relationship should exist between the extent of deviation from the direct path (and hence the length of the path traversed) and time required for a given reaction is an obvious one. Such a direct rectilinear relationship was found in examining the results for Task A. The relationship in the case of Task B proved, however, to be curvilinear with reactions deviating about half way toward the reversion-path consuming the longest times. The general trend of the relationship for the critical stimuli in Task B can be shown by plotting the mean reaction-times against the extents of the deviation in the critical direction, as is done in Fig. 4. (That the curves of this figure are truly representative is confirmed by analysis of the results for each of the individual subjects.) The correlation-ratios between the two variables are relatively low.[11]

The clue to an understanding of this curvilinear relationship lies in the identification of the reactions characterized by

[11] X, reaction-time variable; Y, deviation-variable.

Stimulus 3: $n_{xy} = .49 \pm .04$; $n_{yx} = .40 \pm .05$
Stimulus 8: $n_{xy} = .48 \pm .04$; $n_{yx} = .34 \pm .05$.

the different times. The findings of an analysis of the results are, in brief, that the shortest reaction-times were characteristic of both the normal, correct reactions and the complete, overt reversions—the cases showing most clearly the absence, on the one hand, and the presence, on the other hand, of reversion-influence. For both of these types of reaction the mean reaction-time was the same, 1.5 seconds, and the extreme drop at the two ends of the curves of Fig. 4 is produced

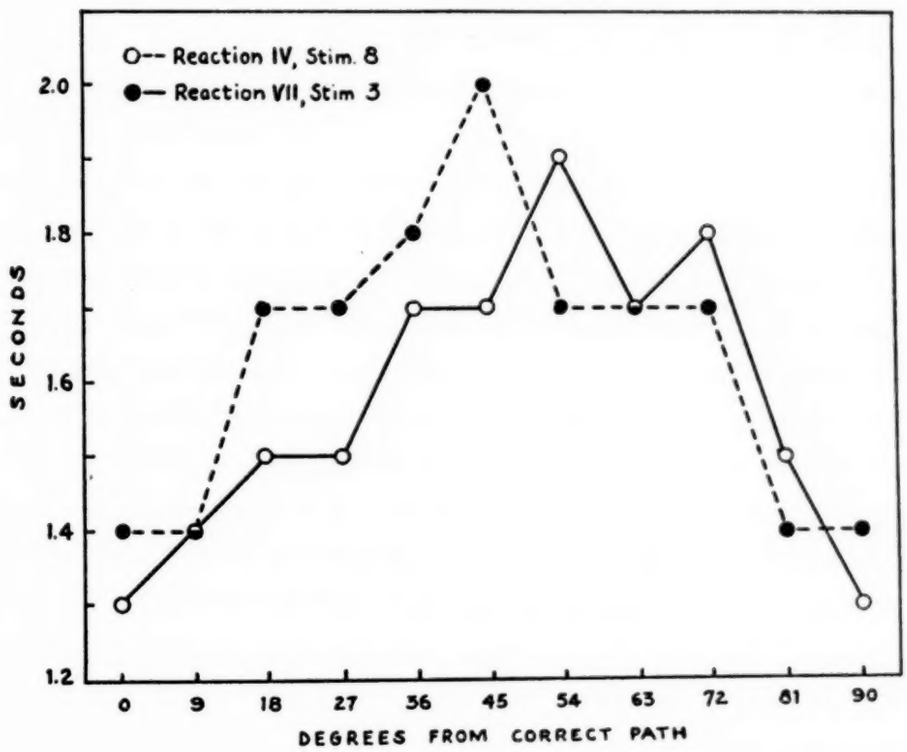


FIG. 4. Mean reaction-times for reaction-movements involving various extents of deviation toward reversion-slots located at 72°.

by the presence at those points of reactions of these two types. The longest times were characteristic of reactions involving partial reversion-errors and falling in the intermediate region of the curves. The mean reaction-time for all reactions showing significant partial errors was 1.8 seconds. It might be expected that the longest times would have occurred naturally in connection with partial errors deviating to the widest extent and falling just to the left of 72° on the curves. The peak actually appears nearer the center of the curves as a result of influences (such as the greater absolute speed associated with

wide-swinging movements) working against the factor of mere length of the path traversed.

These findings point to the conclusion that reaction-time alone is not a direct index of the strength of reversion-influence upon any given reaction. A true reversion-movement may have either short or long reaction-time in dependence upon whether the reversion produces a complete or a partial error. Reaction time, as a gross measure applied to a whole series of reactions in a second task (as is the common practice), may, in certain favorable instances in which partial errors predominate, reflect reversion-influence as an increase in time required.

The preceding analysis has been made in terms of the separate reaction-movements for all subjects. Information concerning the relation of the two measures when applied to the results for individual subjects is available from a comparison for each subject of the frequency of reversion errors with reaction-time (in relation to normal speed in Task A). The results show a relatively low positive correlation ($r = .39 \pm .10$) between the two measures. That even this low correlation-figure does not picture adequately the actual relationship of the two measures as applied to particular individuals is evidenced by the facts that among the subjects showing least increase in reaction-time were three with the very highest percentages (90 percent to 100 percent) of reversions, and that among subjects who showed a decrease in time (positive transfer-effect) were three with more than half (60 percent) of their reactions involving reversion-errors. That the time-scores were primarily influenced by the proportion of partial errors is shown by the fact that almost all subjects showed an increase in reaction-time for stimulus 8 (for which partial errors greatly outnumbered complete errors) whereas only half of the subjects showed an increase for stimulus 3 (where partial errors were less predominant). Throughout the analysis of the individual results the direct measure of reversion-errors showed itself to be more sensitive in revealing specific reversion-effects than was the time-measure.

Of final interest are the figures obtained by calculating for this experiment the gross measures of transfer-effect. Calculation of the transfer-effect in terms of the mean reaction-time for all 50 reactions of Task B in comparison to the mean reaction-time for all 75 reactions of Task A gives a result of zero transfer since the means were identical (1.5 sec.) for the two tasks. Calculation of the transfer-effect in terms of merely the reactions to the critical stimuli gives the result − 7 percent, a low index of negative transfer-effect in consideration of the fact that more than half of the reactions included were previously shown to constitute reversion-errors. (It is of interest to note that the amount of negative transfer-effect for the non-critical stimuli 4 and 7 was higher—13 percent.) If, for the reactions to critical stimuli, extents of deviation (reversion-error) were to be substituted for the time-figures, the percentage indicating negative transfer-effect would reach 383, an absurd value from the point of view of the range of results conventionally obtained by this method of calculation. It is obvious that reaction-time is, in the case of this experiment at least, merely a gross product reflecting the influences of various factors combined in such fashion that important reversion-effects may be almost completely masked.

Summary

On the assumption that phenomena of habit-interference might be more adequately understood through direct study of reversion-errors than by measurement of gross transfer-effects, an attempt was made to devise an apparatus and a technique appropriate to the recording and measuring of reversion-errors. It was considered essential that implicit or partial errors, as well as overt or complete reversions, be detected and subjected to measurement. Code-learning involving simple manual habits was employed, and a direct mechanical means for recording literally the paths of reaction-movements was provided.

Application of the technique in a simple experimental situation requiring change of habit demonstrated clearly the

adequacy of the methods for recording and measuring reversion-errors. The sensitivity of the technique was such as to reveal reversion-influence, predominantly manifested as partial errors, in more than half of the critical reaction-movements. The importance of including partial errors in measuring reversion was demonstrated by the fact that, under the conditions of this experiment which actually favored overt reversions, more than two-thirds of all the reversion-errors produced were of the partial type. Furthermore, results for different reaction-items indicated that the greater the frequency of partial errors, the lower the frequency of overt reversions. This finding suggests a complementary relationship between the two kinds of errors and makes questionable the common practice of inferring the strength of reversion-influence from the frequency of overt reversions alone. A supplementary point of both theoretical and methodological importance was the finding that reversion-influence may extend beyond the critical reactions (changed in the second task) to other closely related reactions and may produce upon such non-critical items effects as great as those upon the critical reactions themselves.

Comparison of the direct measures of reversion (in terms of extent and frequency of errors) with indirect measures based upon reaction-time showed only low positive correlation between results obtained by the two types of measurement. For the separate reaction-movements the relation between extents of reversion-error and reaction-time was definitely a curvilinear one with the longest times occurring in connection with partial errors and with the shortest times equally characteristic of both complete overt reversions and normal, correct reactions. In all of the comparisons made, the reaction-time measures showed themselves inferior in detecting the reversion-phenomena present and less sensitive in revealing the magnitude of reversion-influence. Calculation for the results of this experiment of the usual gross measure of transfer-effect in terms of reaction-time showed zero, rather than negative, transfer-effect; calculation based upon only the critical reac-

tions, more than half of which involved definite reversion-errors, gave the low figure — 7 percent indicating a negligible amount of reversion-influence. It is obvious from these comparisons that only under certain favorable circumstances will reaction-time reveal the presence of reversion-phenomena, and that the adequacy of methods for measuring the strength of reversion-influence by reaction-time is highly questionable.

(Manuscript received July 8, 1939)

THE CONCEPTION OF THE TRUE PATH AND EFFICIENCY IN MAZE LEARNING *

BY A. Q. SARTAIN

Southern Methodist University

It is a matter of common observation that if one learns a stylus maze with vision excluded he usually builds up some sort of a conception of the true path of the maze. It might be supposed, furthermore, that one learns in proportion to the clearness and accuracy of this conception. The present study is an attempt to test the validity of that supposition. The study will raise such questions as the accuracy of the conception after learning is complete; how the conception functions in the process of learning; and whether a conception arrived at as a result of map study is as valuable as an equally accurate one arising from actual experience in the maze.

Materials Used.—The stylus maze used in this experiment was similar to one previously used by Carr.[1] A metal stylus was inserted in grooves in a metal plate in such fashion that the stylus would slide freely except when prevented by stops so located that they could not be discovered visually. The base of the maze was made of a sheet of duralium, twelve inches square and one-half inch thick. To this were bolted twenty-five squares of the same metal, each one and seven-eighths inches square on top, with the under side, or side next to the duralium plate, one and three-eighths inches square. This left a notch or groove one-fourth inch high and one-half inch wide, running completely around the bottom side of each square. These metal squares were arranged in rows of five each, with one-fourth inch between them. Details of the construction are shown in Fig. 1 and the general pattern in

* From the Psychological Laboratory of The University of Chicago. The writer wishes to express his indebtedness to Prof. H. A. Carr, under whose direction this research was carried out.

[1] Carr, H., The influence of visual guidance in maze learning, *J. Exper. Psychol.*, 1921, 4, 399–417.

Fig. 2. The stops in the pathway were made by bolts which came up through the metal plate within the grooves mentioned above. They were placed under the edges of the squares of metal so that they could not be seen. The stylus was shaped like an inverted *T*, and would not pass the stops.

One of the most difficult and important problems involved in the study was that of testing the subject's conception of the true path. The method finally adopted was the use of a questionnaire concerning the correct pathway, the questions being so arranged that they could be answered by merely checking in the appropriate column, 'Yes,' 'No,' or 'Don't Know.' The subject was furnished with a map of the maze *with no pathway shown on it*, but with the spaces between the

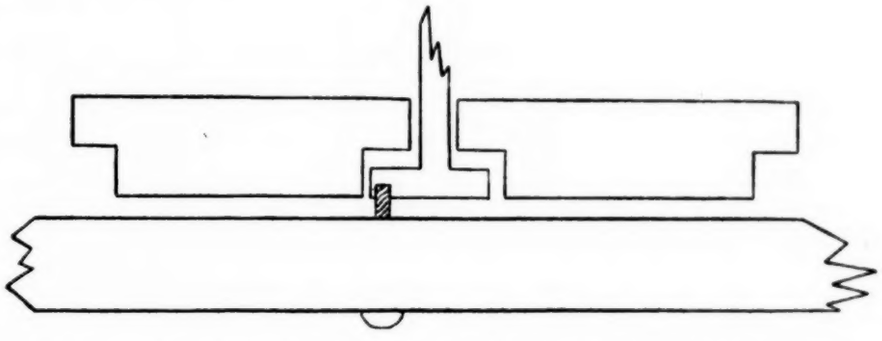


FIG. 1. Cross section through maze, showing base, stylus, metal squares, and stop. (Actual size.)

corners of the metal squares all numbered. These numbers began with '1' in the upper left-hand corner and proceeded from left to right across the map, with a number at each intersection. Thus, across the top there were the numbers from 1 to 6 (with '6' in the upper right-hand corner). Just under the '1' was a '7,' under the '2' an '8,' and so on, until in the lower right-hand corner a '36' was placed. The Questionnaire, based on this map and employing these numbers, was composed of sixty questions concerning the correct pathway. Thirty-six of these were correctly answered by checking 'Yes,' and twenty-four by checking 'No.' Of the false questions several represented merely the wrong direction on the correct pathway. Figure 2 shows the general appearance of the maze, as well as the correct pathway, the blind alleys, and the location of the stops.

In order to determine the sensitiveness of the Questionnaire—that is, its ability to differentiate well between varying degrees of knowledge of the true path—it was administered to twelve subjects before the experiment proper was begun. Accuracy on the Questionnaire, after it had been modified

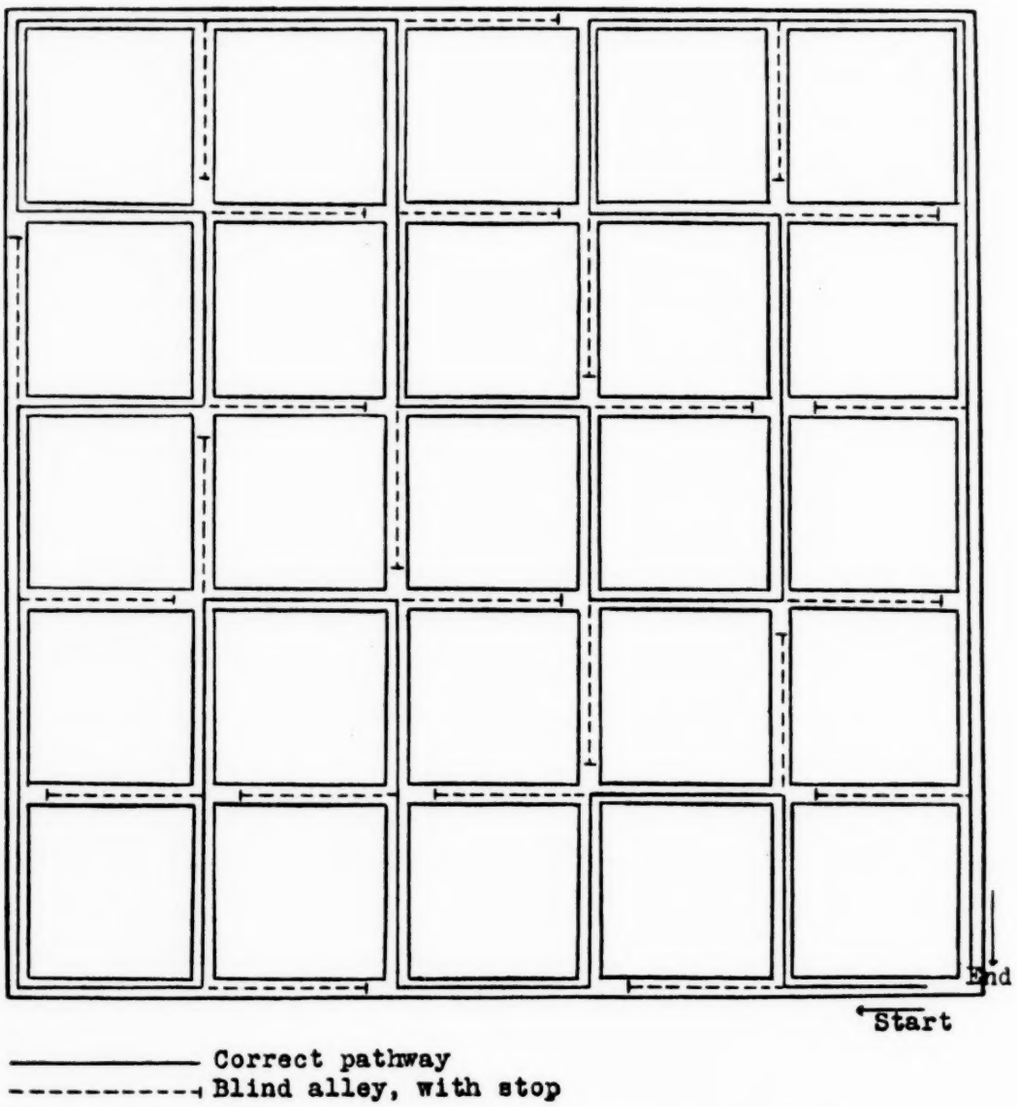


FIG. 2. Map of maze, showing correct pathway, blind alleys, and stops.

somewhat, was found to agree closely with the subject's performance and, in the end, his own estimate of his knowledge. These preliminary trials and subsequent data from the experiment proper seem to justify the conclusion that the Questionnaire served its purpose well. A copy of the Questionnaire is to be found in the Appendix.

Subjects and Conditions of the Experiment.—Six groups of subjects were employed in this study. All of the groups learned the maze with vision excluded by means of a screen. All groups also received typewritten instructions, and were allowed to study the map which showed the general nature of the maze but which had no pathway indicated upon it. They were also allowed thirty seconds of visual inspection of the maze itself. One minute thereafter learning was begun, and it continued until the subject had learned to the extent of three successive errorless trials. An error was considered to be any complete entry into a blind alley or any retracing of the correct pathway by as much as the length of one of the squares.

The maze was learned under three different conditions: first, with no preliminary information concerning the correct pathway but with the Questionnaire to be answered at the end of each third trial during learning (Group I); second, with no preliminary information concerning the correct pathway and the Questionnaire to be filled out only at the end of the learning period (Group II); and third, with varying amounts of preliminary study of a map showing the correct pathway (Fig. 2), the Questionnaire being filled out (1) after the map study and before attempting the maze and (2) after the learning was completed. In this third condition one group studied the map showing the correct pathway for fifteen seconds (Group IIIa); another studied it for thirty seconds (Group IIIb); a third, for sixty seconds (Group IIIc); and a fourth, for one hundred twenty seconds (Group IIId). Thus, for all four of these last named groups there were obtained, by means of the Questionnaire, an Initial Accuracy score (from study of the map showing the correct pathway but without any trials in the maze) and a Final Accuracy score (at the end of learning). As already implied, Group I yields a series of Intermediate Accuracy scores (after each third trial) and a Final Accuracy score. The second group has only the Final score. For all six groups records were kept of time for learning, errors made, and trials to learn. Fifty college students, mostly sophomores, comprised Group I. There were thirty comparable subjects in each of the other five groups.

History of the Problem.—There seem to be no previous studies of the exact questions involved in this experiment. Indeed the number of relevant studies is quite small. The question of the type of imagery used in maze learning and how that imagery functions in the learning process was dealt with by Perrin,[2] who reports that he was 'unable to make any correlation between the type or combination of image processes used with efficiency in maze learning.' Furthermore, scores on tests of memory showed no relationship to efficiency in learning. Illusions as to the correct pathway were general, but as a rule they did not seem to produce bad results. On the whole there was slight correlation between efficiency in learning and 'pure trial and error,' 'pure reasoning,' 'ideational learning,' memory, or imagery.

On the other hand, Warden [3] found evidence that there are three methods of learning a stylus maze, word reaction, visual imagery, and kinesthesis. He found that those who used the first method were greatly superior to those employing the third, with those using the second method intermediate in efficiency.

Lambert and Ewert [4] confirmed the findings of Warden. In addition, they discovered that the more complete the instructions, the less of the motor and the more of the verbal methods were used. Better reproduction of pattern was also displayed by the verbal group.

Husband [5] also obtained results agreeing with those of Warden. He found that the verbalizers were much more slow and deliberate than those using the motor method, but that in the end the former showed a superiority in total time to learn, a superiority that increased with the difficulty of the task.

[2] Perrin, F. A. C., An experimental and introspective study of the human learning process in the maze, *Psychol. Monog.*, 1914, No. 70, 1–97.

[3] Warden, C. J., The relative economy of various modes of attack in the mastery of a stylus maze, *J. exper. Psychol.*, 1924, **7**, 243–275.

[4] Lambert, J. F., and P. H. Ewert, Part I: The effect of verbal instructions upon stylus maze learning, *J. gen. Psychol.*, 1932, **6**, 377–399.

[5] Husband, R. W., Analysis of methods in human learning, *J. genet. Psychol.*, 1931, **39**, 257–278.

Scott [6] also came to the conclusion that verbal imagery was most efficient in maze learning and that there was a 'pronounced tendency to change from motor to verbal learning' during the learning process. He also secured evidence that the maze pattern is retained in memory and at times 'does function in the relearning of the same maze or a similar maze,' though often 'nonconsciously.'

On the other hand, Sackett,[7] attempting to discover the relationship between the retention of a maze habit and the amount of symbolic rehearsal, found a large majority of his subjects used the visual method, with the verbal method next, and finally the motor method. Though only eleven percent of the subjects used one method exclusively in learning, forty-one percent did so in rehearsal and even more in relearning. The visual method, however, predominated in all cases.

Finally, Scott [8] attempted to reconcile the discrepancies in the studies of Perrin, Warden, Sackett, and the others on the assumption that the sort of image secured in maze learning depends upon the type of maze pattern employed. He held that a systematic maze [9] 'is found to call forth in the majority of cases a certain type of imagery peculiar to it,' while an unsystematic [10] one produces variation from subject to subject and within the same subject.

Thus we see that it is generally agreed that, in the process of learning a stylus maze, at least a majority of subjects build up some sort of a conception of the maze pattern. At times this conception may be visual, at other times it may be verbal. In still other cases the learning may be a motor accomplishment. In the present study, therefore, we are making no

[6] Scott, T. C., The retention and recognition of patterns in maze learning, *J. exper. Psychol.*, 1930, **13**, 164–207.

[7] Sackett, R. S., The relationship between amount of symbolic rehearsal and retention of a maze habit, *J. gen. Psychol.*, 1935, **13**, 113–130.

[8] Scott, T. C., Imagery—to what extent is it a function of the maze pattern? *J. genet. Psychol.*, 1936, **49**, 271–278.

[9] By a systematic maze Scott means one composed of a number of similar parts or of units—to use his own words, 'reducible to some common unit,' 'reducible to a unit of construction.'

[10] *Ibid.*

assumptions as to the experiential nature of the conception possessed by the subject, for we are concerned only with its accuracy as related to efficiency of learning.

Another problem involved in this experiment is the effect of visual guidance on maze learning. In this connection Carr,[11] using an experimental set-up similar to the one employed for Group III of this study except that no questionnaire was used, found that visual guidance was effective, especially in the prevention of errors, although the effectiveness was not proportional to the amount of guidance. He also found that, after a certain point was reached, further visual guidance was not effective, since after a time all that may be learned visually has been learned. Furthermore, some subjects who learned with vision excluded made errors when required, after learning, to trace the pathway with vision.

Carr used, for visual guidance, maps revealing varying amounts of information about the correct pathway. He found that "the effectiveness of the guidance upon trials and errors was proportional to the knowledge obtained from the diagram." Looking at the maze was more effective than merely looking at a map of the maze, when no pathway was shown on the map.

When the subjects were required during learning to identify the correct pathway on a map they made many errors at first. Even after the pathway had been correctly identified (after an average of 3.8 trials and 24.2 errors), to traverse the pathway correctly required considerable additional effort (8.0 additional trials, 1.7 additional errors). Conspicuous differences between subjects were noted in this regard, as in other phases of the experiment. Among the author's conclusions one finds this statement:

> Our results at least indicate the probability that the visual modes of control are much more effective in the acquisition and performance of acts of skill than hitherto suspected.[12]

Carr and Osbourn[13] in a similar experiment arrived at

[11] Carr, *op. cit.*

[12] *Ibid.*, p. 417.

[13] Carr, H. A., and E. B. Osbourn, Influences of vision in acquiring skill, *J. exper. Psychol.*, 1922, 5, 301–311.

substantially the same conclusions. They too found that while visual study of a maze is effective in learning a maze, it is not effective in exact proportion to its amount; and further, that "the mastery of the maze in visual terms does not necessarily enable the subject to traverse it successfully with vision excluded." Likewise they found that for a majority of their subjects the introduction of vision disturbed the maze performance if learning up to that point had been carried on with vision excluded. Again, however, the fact of individual differences is emphasized.

Results of the Experiment.—The findings of the present study may be presented under three main headings: first, the relation of Final Accuracy on the Questionnaire to errors, time, and trials for learning; second, the relation of Intermediate Accuracy on the Questionnaire to errors and time up to the particular trial; and third, the relation of Initial Accuracy on the Questionnaire to errors, time, and trials for learning. The first of these relationships will show whether faster learners tend to build up a more accurate knowledge of the pathway. The second will deal with the growth in the accuracy of the conception and with the relationship of the increase in accuracy to learning at various points in the process. The third will show the effect upon subsequent efficiency of the conception obtaining in the beginning of actual maze learning.

One noteworthy fact concerning Final Accuracy is the high score made when the Questionnaire is filled out after learning. Data on this point are given in part in Table 4. Of fifty-nine possible correct answers, for example, Group I had a mean score of 56.28 correct responses (with a probable error of the mean of only .34); Group II had a mean of 53.33 correct responses (probable error, .75); Group IIIa, 56.73 correct answers (probable error, .31); Group IIIb, 56.73 correct answers (probable error, .26); Group IIIc, 57.43 correct answers (probable error, .39); and Group IIId, 57.00 correct responses (probable error, .27).

Thus it may be seen that, as a general rule, subjects tend to have an accurate conception of the true path by the time they reach the standard required for the last three trials,

although a perfect conception is obviously not necessary for the requisite final performance. The small probable errors indicate a strong tendency for the scores to bunch at the upper end of the distribution, a fact which will be referred to later in another connection.

The relationship of Final Accuracy on the Questionnaire to total time, errors, and trials is shown in Table 1. Generally

TABLE 1

RELATION OF FINAL ACCURACY ON QUESTIONNAIRE TO RATE OF LEARNING

Group	Coefficient of Correlation between		
	Accuracy and Total Time	Accuracy and Total Errors	Accuracy and Trials
I	−.1606	−.3069	−.5601
II	−.5016	−.4461	−.3934
III Total	−.2771	−.2021	−.2417
IIIa	+.0406	+.0757	+.0187
IIIb	−.4901	−.3774	−.4426
IIIc	−.3373	−.2653	−.1530
IIId	−.3504	−.3026	−.3197

speaking, the correlations are quite low, although in each case except for Group IIIa the correlations are all negative, that is, a higher score on the Questionnaire tends to go with lower error, time, and trial scores. In other words, the more accurate the conception, the faster is the rate of learning. For Group II all coefficients are −.39 or above, but the average coefficient for all groups is only −.28. Thus it may be seen that while there is an inverse relationship between Final Accuracy and time, trials, and errors, the relationship is relatively low for all groups. With one exception correlations are higher for Group II than for Groups I and III but not markedly so.

The small size of the average correlation in this connection is somewhat surprising. In part the lowness of the figures doubtless is due to the distribution of the Final Accuracy scores. As has already been stated, there is a tendency for those scores to gravitate towards the upper part of the distribution, with the consequent effect, of course, of lowering the coefficient of correlation. It seems quite likely, moreover,

that the greater distribution of scores for Group II helps to explain its somewhat higher coefficients of correlation. However, in spite of these facts the conclusion seems quite clear: the relationship between the conception held by an individual and his rate of learning, while definitely present, is low. It is worth noting that for the fifteen second group (Group IIIa), as has already been pointed out, the correlation is non-existent. Perhaps fifteen seconds is too short a time to influence the final conception, but, as will be seen later, it has a definite effect on efficiency in learning.

It should also be noted that the Final Accuracy scores for Groups I, IIIa, IIIb, IIIc, and IIId are decidedly higher than those for Group II. It will be recalled that Group I filled out the Questionnaire after each third trial and that Groups IIIa, IIIb, IIIc, and IIId all studied maps of the maze. Evidently these activities tend to improve Final Accuracy.

The second problem raised by the study concerns the development of the conception and its relation to learning. As one would expect, there are here as elsewhere great individual differences. A few of the subjects are never able to fill out the Questionnaire with any accuracy, even at the conclusion of the experiment, nor do they show evidence of any clear conception of the true path at any time during learning. Other subjects, on the contrary, seem to have the conception before they can follow the path in the maze. Most subjects, of course, lie between these extremes.

That the conception grows very much from the beginning to the end of the experiment is at once evident when the individual records are inspected, but the data are rather difficult to present in tabular form because of the fact that as soon as a subject completed three successive errorless trials he dropped out of the group. One subject, for example, finished with the sixth trial, another with the ninth, two more with the tenth, etc. Thus, the average Final Accuracy scores after each third trial would be based on fewer and fewer subjects, and these subjects are the slowest learners. However, the average Accuracy score of all subjects after the third trial in the maze was 34.2 correct answers, or nearly sixty percent of a perfect

score. And since the average Final Accuracy score for Group I was approximately 56.3, as already pointed out, this indicates an average gain of 22.1 correct answers during the course of the experiment. Furthermore, the highest scores were made at or near the end of the learning period. Thus, for 19 of the 50 subjects the highest accuracy attained during the experiment was just after the completion of learning, and 43 of the 50 made their highest score within three trials of the last.

The conclusion from this section of the study, therefore, substantiates one previously drawn: increasing ability to execute a successful maze performance is accompanied to some extent by an increasing accuracy of the conception of the true path of the maze.

Correlations between Intermediate Accuracy scores and learning and between Intermediate Accuracy scores and performance are shown in Tables 2 and 3. Table 2 shows the correlations between Intermediate Accuracy (that is, accuracy at the end of the third trial, sixth trial, ninth trial, etc.) and time consumed and errors made to that point in the learning process. As previously noted, because of the dropping out of subjects after three successive errorless trials, the correlations are based on fewer and fewer subjects as one reads down the table. Correlations obtained in this way are considerably higher than those of Table 1, if the smaller groups are disregarded. Indeed the average coefficient through the fif-

TABLE 2

RELATION BETWEEN ACCURACY ON QUESTIONNAIRE AND LEARNING TO THAT POINT

Trial	Coefficient of Correlation between Correct Answers and		Number of Subjects
	Total Time (to that Point)	Total Errors (to that Point)	
3	−.4780	−.2940	50
6	−.4100	−.6010	50
9	−.5773	−.6440	49
12	−.4381	−.6414	45
15	−.3730	−.6959	39
18	−.1462	−.4424	31
21	+.1773	−.0196	22
24	+.3025	+.1300	18
27	−.3285	+.0440	11
30	+.3874	+.1657	8

TABLE 3

RELATION BETWEEN ACCURACY ON QUESTIONNAIRE AND PERFORMANCE ON NEXT TRIAL

Trial	Coefficient of Correlation between Correct Answers and		Number of Subjects
	Total Time (on Next Trial)	Total Errors (on Next Trial)	
3	−.2763	−.2830	50
6	−.3314	−.4151	49
9	−.3612	−.4880	48
12	−.2089	−.5874	43
15	−.0717	−.4517	37
18	−.4089	−.3677	27
21	−.3933	−.5370	21
24	−.4934	−.5581	14
27	−.2511	−.3573	9
30	−.0828	+.0229	7

teenth trial in Table 2 is −.52. This may mean that the better learners develop more accurate conceptions and drop out earlier while the poor learners reduce the ultimate correlation, or it may be due to the fact that the range of the accuracy scores is greater. At any rate the developing conception correlates more highly with learning scores to that point than the final conception does with final learning scores. It is possible, therefore, that while the development of a conception of the pathway of the maze is by no means the only factor in learning, it is a factor of some importance.

Another aspect of the experiment, previously considered in part, is the effect on Final Accuracy, time, errors, and trials of filling out the Questionnaire at the end of each third trial. The answer to this question may be found by comparing the scores for Groups I and II, as these appear in Table 4. For all four criteria the advantage is with those who answered the Questionnaire after each third trial. Thus, these subjects finally answered, on the average, 56.28 questions correctly, while those who answered the Questionnaire only at the end of the experiment replied correctly on the average to only 53.33 questions. Group I likewise made only 508.08 errors per subject, while Group II made 617.23 errors. Group I required only 20.86 trials for learning, Group II, 27.70 trials. Group I consumed on the average 2312.02 seconds in learning,

while Group II required 2547.70 seconds. The superiority of Group I may be explained in part on the basis of motivation, for the more quickly the subjects could finish, the fewer times the rather long Questionnaire had to be filled out. But this seems hardly sufficient to explain differences so large. The subjects frequently reported that an opportunity to 'collect their thoughts' about the correct pathway was definitely helpful, and certainly the evidence cited above seems to bear them out.

When one observes the correlations between Intermediate Accuracy on successive third trials and time and error scores on the next succeeding trials (Table 3), he is struck by the fact that they are usually lower than the ones just dealt with. Doubtless the chance factors influencing the scores for a single trial in the maze are relevant here. Why these correlations should reverse the trend and actually be higher during the last few trials is not clear.

A noteworthy feature in this connection is the effect of filling out the Questionnaire on scores on the next succeeding trial. Here, on the average, time and error scores practically always rose over the trial preceding, that is, more time was required and more errors made on the fourth trial than on the third, on the seventh than on the sixth, etc. It is suggested by way of explanation that erroneous conceptions were worked out in answering the Questionnaire and that these conceptions induced errors when tried out. However, experience with the maze and the blank map frequently gave the subject a better conception of the correct pathway when he had discovered his errors. This conclusion is in keeping with statements frequently made by subjects on the first and second trials following the Questionnaire.

Thus far we have seen that the conception held by the subject correlates, sometimes fairly well and at other times less well, with his rate of learning in the maze. It has not been possible to show a causal relationship, however, for it is entirely possible that the conception merely accompanies the learning, without exerting any influence upon it. It now remains to consider the effects on learning of a conception

induced before experience with the maze itself (Groups IIIa, IIIb, IIIc, and IIId).

That map study definitely induces a conception of the pathway is apparent from Table 4. Fifteen seconds of study, for example, enabled the average subject to answer more than two-thirds of the questions correctly, and one hundred twenty seconds raised this to almost nine-tenths. And though the relationship between time and Initial Accuracy is by no means one-to-one, the group studying the map for thirty seconds stands next to first, and the one studying for sixty seconds has the next to the best score in Initial Accuracy.

TABLE 4
EFFECT OF MAP STUDY

Group	Duration of Map Study	Initial Accuracy on Questionnaire	Final Accuracy on Questionnaire	Trials	Errors	Time
I	—	—	56.28	20.86	508.08	2313.02
II	—	—	53.33	27.70	617.23	2547.70
III Total		47.05	56.97	13.03	148.16	1041.43
IIIa	15 sec.	40.96	56.73	16.93	209.53	1341.31
IIIb	30 sec.	43.06	56.73	16.00	249.46	1471.40
IIIc	60 sec.	51.50	57.43	10.40	87.73	840.75
IIId	120 sec.	52.67	57.00	8.80	45.93	539.26

One of the problems in this connection is the effect of such a conception on trials, errors, and time for learning. Table 4 brings out the essential relevant facts. In all three cases the groups studying the map were definitely superior, the superiority increasing with increasing amounts of map study. It is true that differences between Groups IIIa and IIIb are slight, and in the case of errors and time actually favor the group studying for the shorter period. But such is certainly not the case for Groups IIIc and IIId. In these cases the superiority of longer map study is clearly evident.

Group IIId is likewise superior to Group IIIc. Although differences in Initial Accuracy are small, for errors and time especially Group IIId is far ahead.

It is interesting to note that with all the success of groups spending more time in map study, the superiority, with the possible exception of errors, is not proportional to the amount

of time spent in learning. Indeed, eight times as much time in study (Groups IIIa and IIId) lessened trials by only one-half and time by less than two-thirds. Errors, however, are diminished by almost five-sixths. Similar results are obtained by comparisons of other groups. Errors, therefore, seem to be affected most.

Inspection of Table 4 leads one to conclude that for this particular problem, when time intervals are short, additional time makes little difference (Groups IIIa and IIIb). This conclusion, however, does not obtain for the longer intervals. Had still longer periods been used additional time might have had less effect.

Another approach to the problem of the relation of the initial conception to time, errors, and trials for learning is to correlate Initial Accuracy scores with these learning criteria. Here, as Table 5 reveals, the correlation is fairly high, the

TABLE 5

RELATION OF INITIAL ACCURACY ON QUESTIONNAIRE TO LEARNING

Group	Duration of Map Study	Coefficient of Correlation between Correct Answers and		
		Total Time to Learn	Total Errors to Learn	Trials to Learn
IIIa	15 sec.	−.6584	−.4731	−.3839
IIIb	30 sec.	−.3533	−.2005	−.4540
IIIc	60 sec.	−.6960	−.7059	−.6958
IIId	120 sec.	−.6349	−.7100	−.5837
III Total		−.6351	−.4975	−.6354

coefficients rather considerably exceeding those obtained elsewhere in the experiment. Only one coefficient is below −.35, and the average is −.55. The size of these correlations may best be appreciated when they are compared with those of Table 1, dealing with Final Accuracy as it relates to these measures. With a single exception the former are higher and in most cases markedly so.

Thus, because the more accurate the initial conception, the lower the time, errors, and trials required for learning and because Initial Accuracy correlates definitely with the learning criteria, we arrive at the conclusion that the conception held at the beginning of maze learning may definitely influence the rate of learning. And certainly if this statement holds for

Groups IIIa, IIIb, IIIc, and IIId, it may safely be said to obtain for Groups I and II also. It will be remembered that for the latter groups the only data available showed a correlation of conception and rate of learning, but no method of showing actual influence was available. Data just presented for Groups IIIa, IIIb, IIIc, and IIId remedy that defect.

Data are also available on another problem having to do with the effect of the conception on learning: is a certain accuracy of conception gained through map study as effective as an equally accurate conception built up in actual experience in learning the maze path? To answer this question Groups IIIa, IIIb, IIIc, and IIId were compared with subjects taken from Group I, those subjects being selected in each case whose average Intermediate Accuracy was practically the same as the Initial Accuracy of the particular group concerned. To illustrate: the Initial Accuracy of Group IIIa was 40.96 correct answers. When all subjects from Group I who at the end of any third trial had an Intermediate Accuracy score of 40, 41, or 42 were considered as a sub-group (S–Ia), it was found that their average accuracy score was 41.00. The question was then to determine whether Group IIIa learned the maze with less errors, time, and trials than did this sub-group, assuming that the next trial after the 40, 41, or 42 score was the *first trial* for the subjects in the sub-group. The sub-group to be compared with Group IIIb (Initial Accuracy 43.06) was composed of those subjects making at some time Intermediate Accuracy scores of 42, 43, or 44 (S–Ib). Group IIIc (Initial Accuracy 51.50) was compared with those scoring 51 and 52 (S–Ic), and Group IIId (Initial Accuracy 52.67), with those scoring 52 and 53 (S–Id). The results obtained are shown in Table 6.

As far as the learning criteria are concerned, the results show clear superiority for the sub-groups in every comparison. Thus, Group IIIa required 16.93 trials to learn, while Group S–Ia required only 14.09 trials. Group IIIa required 209.53 errors as against 82.14 errors for Group S–Ia, and 1341.31 seconds as against 700.61 seconds. Similar comparisons could be made for the other groups, with the greatest differences in errors in all cases and the least in trials. It is

TABLE 6

EFFECTIVENESS OF CONCEPTION GAINED THROUGH MAP STUDY AS COMPARED TO THAT GAINED THROUGH EXPERIENCE IN MAZE

Group	Number of Subjects	Initial Accuracy on Questionnaire	Final Accuracy on Questionnaire	Trials	Errors	Time
IIIa	30	40.96	56.73	16.93	209.53	1341.31
S-Ia	22	41.00	55.23	14.09	82.14	700.61
IIIb	30	43.06	56.73	16.00	249.46	1471.40
S-Ib	23	43.13	55.44	11.78	50.70	567.53
IIIc	30	51.50	57.43	10.40	87.73	840.75
S-Ic	18	51.28	55.78	7.39	12.83	299.70
IIId	30	52.67	57.00	8.80	45.93	539.26
S-Id	23	52.57	56.70	5.70	5.52	229.29

clear, therefore, that the initial trials of the sub-groups influenced learning in other ways than simply in building up the conception.

To guard against the possibility that the better record of the sub-groups might be attributed to the possession of greater learning ability, the average trial, time, error, and Final Accuracy scores of each sub-group *for the entire period of its learning* were compared with those of Group I as a whole. It was found that three of the four sub-groups were actually below the average in learning ability, and the fourth but slightly above. The results noted, then, could hardly be attributed to greater ability.

The Final Accuracy scores of Groups IIIa, IIIb, IIIc, and IIId deserve some attention. In the first place, these groups definitely excelled the corresponding sub-groups in Final Accuracy (Table 5), although for the group studying the map the longest time (IIId) the difference is slight. That a difference of this sort should occur, however, is quite logical when one considers that Groups IIIa, IIIb, IIIc, and IIId had studied a map on which the pathway was clearly traced out, while Groups S–Ia, S–Ib, S–Ic, and S–Id had acquired their information from the maze itself. Thus, each set of subjects seems to have learned best what it practiced most, the map-study groups to recognize the pathway in terms of a map and the sub-groups to traverse more efficiently the true pathway.

In the second place, there is but little difference in Final Accuracy between Groups IIIa, IIIb, IIIc, and IIId (Table 4). Groups IIIa and IIIb were lowest with exactly the same score, Group IIId was next, and Group IIIc highest. As is evident, however, from a consideration of the means and probable errors, none of the differences are statistically significant.

In the third place, when Groups IIIa, IIIb, IIIc, and IIId are compared with Group I, differences in Final Accuracy are again insignificant. But assuming that a difference between the means is significant when it is four times the probable error of the difference, we find that Groups IIIa, IIIb, IIIc, and IIId are superior to Group II. This, of course, would be expected in view of the fact that Group II had only one contact with the map, and therefore devoted practically its entire energy to learning to traverse the maze and not to answering the Questionnaire.

Parenthetically, the difference between the means of Groups I and II is 3.7 times the probable error of the difference. Since Group I also had several contacts with the Questionnaire, the near-significance of this difference supports the conclusion just drawn.

Finally, from a comparison of the coefficients of correlation of time, errors, and trials with Final and Initial Accuracy it is at once apparent that Initial Accuracy correlates more highly than does Final Accuracy. As previously noted, the greater bunching of the latter scores towards the upper end of the distribution likely helps explain this difference. It is entirely possible, however, that the relationship between performance and accuracy is higher at an earlier stage in the learning process than it is at the end. Indeed the data of Table 2 lend some support to this possibility.

Summary and Conclusions.—Two hundred college students, divided into six groups, one of fifty subjects and the other five of thirty each, learned a visual type of maze with vision excluded. From time to time in the experiment and for all subjects at the end of the experiment the accuracy of the conception of the true path of the maze was tested by means of a questionnaire designed for the purpose. The group

of fifty subjects filled out the Questionnaire at the end of each third trial in the maze, and four of the other five groups filled it out after varying amounts of study of a map of the maze and after learning. One group filled out the Questionnaire at the end of the learning period only.

Several conclusions may be drawn from the data presented. First, each individual does develop a conception of a stylus-maze pattern as he learns the maze. In some individuals this conception seems to be quite faulty and in others perfect or well-nigh so. As was pointed out, however, the conception finally attained a high degree of accuracy on the average.

Second, this conception is extremely inaccurate in the early stages of learning, but it increases in accuracy as learning progresses.

Third, the relationship of the accuracy of the developed conception to efficiency in learning is relatively small. This is true whether one judges by final accuracy or intermediate accuracy but particularly so in the former case. It seems, therefore, that learning a stylus maze is something other than merely building up a conception of the pattern.

Fourth, a conception developed in the maze itself, as a result of running the maze, is much more effective than one developed from map study.

Fifth, the accuracy of the conception gained through map study correlates more closely with the learning criteria than does the final accuracy of the conception. This may be due to an artifact of the data, the distribution of scores in final accuracy being quite restricted, or to the possibility that the conception functions more effectively in the earlier than in the later stages of learning.

Sixth, pauses for working out one's conception on a blank map are of considerable benefit ultimately, even though the pauses prove to be distractive on the next trial.

Finally, to induce the conception artificially by means of map study is definitely of value, although, as already pointed out, knowledge gained from the maze itself is more efficacious. In general the amount gained by map study increases with the amount of study, although there is no reliable difference for the shortest intervals.

This study might be viewed as forging one more link in the chain which restrains and makes untenable that intellectualistic view of psychological functioning which holds that maze learning is proportional to knowing and hardly to anything else. Learning is clearly related to knowing, but the relationship is far from directly proportional.

(Manuscript received May 29, 1939)

APPENDIX

Directions.—By referring to the map before you, answer the questions below, checking in the column opposite the question, 'Yes,' 'No,' or 'Don't Know.' Please note the direction implied in the question; for instance, to say *from 36 to 35* is not the same as to say *from 35 to 36*. The first question is correctly checked as an example.

QUESTIONNAIRE FOR DETERMINING ACCURACY OF THE CONCEPTION OF THE MAZE PATTERN

Should One Go from	Yes	No	Don't Know	Should One Go from	Yes	No	Don't Know
36 to 35?	X			1 to 2?			
28 to 29				19 to 13			
21 to 20				3 to 4			
35 to 34				16 to 15			
32 to 31				8 to 9			
28 to 27				7 to 1			
29 to 28				9 to 15			
20 to 26				16 to 10			
33 to 27				13 to 14			
26 to 20				3 to 9			
31 to 32				23 to 22			
28 to 34				6 to 12			
25 to 19				11 to 17			
27 to 21				4 to 10			
25 to 26				12 to 18			
35 to 29				16 to 22			
22 to 21				4 to 5			
34 to 33				22 to 23			
26 to 32				17 to 11			
31 to 25				5 to 4			
14 to 8				18 to 24			
13 to 19				10 to 4			
14 to 15				11 to 12			
9 to 10				23 to 17			
2 to 3				24 to 18			
7 to 8				5 to 6			
15 to 21				18 to 17			
9 to 3				24 to 30			
8 to 7				11 to 10			
15 to 16				30 to 36			

A STUDY OF CONDITIONING IN THE NEONATE [1]

BY DELOS D. WICKENS AND CAROL WICKENS

The University of Colorado

Problem.—The purpose of the present experiment was to investigate the problem of the conditionability of a withdrawal movement in the newborn infant, and, if positive results were obtained, to study other conditioning phenomena such as experimental extinction and spontaneous recovery in these subjects. This general problem has already been attacked in similar subjects by Marquis (**4**), who used sucking as her unconditioned response, and by Wenger (**8**), who employed lid closure and the withdrawal of the leg to shock in addition to sucking as a response. Although both of these studies obtained results which the writers interpreted as positive, neither of them is entirely conclusive and both have been subjected to criticism. Pratt (**5**) and others (**8**) have suggested that Marquis has not clearly shown that the response to the conditioned stimulus which she obtained after her training series might not have occurred without this training. Further, Wenger, using a procedure similar to that of Marquis, obtained negative results for this particular response. Wenger's own study, although careful and approaching objectivity, must be criticized because of the limited number of subjects used; in many groups only two or three infants were used and one important control group consisted of but four subjects. Since there is a considerable overlapping between the results of his control and his experimental group, this limitation in the number of subjects becomes a serious criticism. A further consideration of this point will be given in discussion of our results.

Because of these objections, it was felt that an additional experiment might clarify the question as to whether or not the newborn infant can become conditioned.

[1] This work was done in the Infant Research Laboratory of the Ohio State University.

Subjects.—Subjects were under-ten-day infants born in the Ohio State University Hospital. Throughout this experiment we are indebted to Dr. Floyd C. Dockeray, Department of Psychology, who made available to us the subjects and facilities of the Ohio State Infant Research Laboratory, and to Dr. Andrew Rogers, Head of the Obstetrics Division of the Ohio State University Hospital. Thirty-six infants in all were used in the actual experiment. These were full term babies about equally divided as to sex. Approximately one third of the subjects were of at least partial negro parentage and the remainder were of white parentage. The infants were from three to five days when they were brought in for the first experimental session, and, the study continuing for four days, the subjects were from seven to nine days of age at the last experimental period.

Apparatus and Procedure.—All experimentation was done with the infants inclosed in the sound proof, thermostatic isolation cabinet, described by Dockeray and Valentine (2).

Before the formal experimentation on the 36 final subjects was begun, various sorts of stimuli were administered to a number of other infants with a view to determining the optimal unconditioned stimulus. From these investigations it was concluded that the best type of stimulus for the study of conditioning of a withdrawal movement was a shock to the sole of the foot. To this stimulus the infant usually responds bilaterally with flexor movements of legs, feet, and toes, often followed by extensor movements. This response is usually more marked in the member receiving the shock than it is in the other member. Shock then was the unconditioned stimulus. This was delivered from a Harvard inductorium, the current being derived from a 6 volt storage battery. The electrode consisted of two 4 × 36 bolt heads attached to the surface of a flexible rubber disc about the size of a penny. This was attached to the sole of the right foot by a strip of Sealtex tape. The conditioned stimulus was the sound of a muffled buzzer, placed in the cabinet about two feet from the infant's head. The buzzer and shock were administered by hand, with the buzzer (in the paired stimulation) preceding the shock by about a fourth of a second and continuing about a fourth of a second more, along with the shock. This hand method of giving stimulations was, of course, necessarily subject to some variability. Stimulations were spaced one, two, three, or four minutes apart according to a varied and irregular pattern. An attempt was made to stimulate the infant only when his activity level corresponded with stages three to five of Wagner's criteria of sleep (7), *i.e.*, when it was neither crying nor highly active and yet not in an extremely deep sleep. Since the babies did not always coöperate, the time intervals between the presentations of the paired stimuli occasionally ran over the one to four minute figures mentioned above. The same infant was always run at the same time on succeeding days, the times being directly following either the two or four o'clock feeding periods. The two experimenters observed and recorded the responses of the infants, one noting the movements of one

limb and one the movements of the other. The responses were recorded upon form sheets in terms of flexor and extensor movements of the leg, foot, and toes.

Three separate groups, consisting of 12 infants each, were used—an experimental group and two controls. For all three groups items of initial procedure were the same. The infant was placed in the isolation cabinet, the electrode having been attached to the sole of the right foot. It was allowed about five minutes to become adjusted to the situation; then tests for responses to the buzzer alone were administered. Three different stimulations of the sound of the buzzer were given, and, if the infant responded by any movement of leg, foot, or toe, he was discarded and not used in any of the three groups. From this point on, the procedure for the different groups varied.

The experimental group received 36 paired stimulations of buzzer and shock, 12 each day for three consecutive days. At the end of the third day, the buzzer alone was sounded to test for the presence of conditioning, and these stimulations of the buzzer were administered until there were three consecutive failures to respond—the criterion of extinction in this experiment. On the fourth day, the buzzer alone was administered until the response was extinguished. This was done, of course, to test for the spontaneous recovery of the previously extinguished conditioned response.

In view of the possibility that positive results in the above group, if present, might be due to the effect of repeated presentations of shock, the unconditioned stimulus, rather than to the pairing of the buzzer and shock, the following control group was set up, which we shall call Control Group 1. Following the usual tests to the buzzer alone, the 12 infants of this group were given 36 shocks alone, 12 each day, for three consecutive days. Thus, they never received a paired stimulation of shock and buzzer. At the end of this period, on the third and fourth days, the buzzer alone was sounded, tests for extinction and spontaneous recovery of the response, if any, being administered just as in the experimental group.

Control Group 2 was run to check the possibility that any responses to the buzzer on the third day might be due solely to the factor of maturation rather than to training. The procedure for this group consisted in testing for responses to the buzzer alone on day 1; then, on day 3, after two days delay, during which time no stimulations were given by the experimenters, administering a second test for responses to the buzzer alone.

Results.—The results for the experimental group are presented in Table I. Each baby is listed separately. Day 3 is,

TABLE I

SHOWING THE NUMBER OF RESPONSES TO THE BUZZER ALONE FOR THE EXPERIMENTAL GROUP ON THE THIRD AND FOURTH DAYS OF THE EXPERIMENT, DAY 3 BEING THE ORIGINAL TEST FOR THE PRESENCE OF CONDITIONING, AND DAY 4 BEING A TEST FOR SPONTANEOUS RECOVERY

Number of Responses	Subjects											
	He	Ca	De	Ch	Re	Mc	Bo	Jo	Ot	Fo	Gr	Ma
Day 3....	2	2	4	0	3	1	5	2	0	1	0	3
Day 4....	8	9	–	0	4	10	7	1	8	1	0	0

of course, the time of the test for presence of the conditioned response, following the training series of 36 paired stimulations; day 4 was the day of tests for spontaneous recovery of the response which had been extinguished on the previous day. By number of responses is meant the number of times the infant responded to the buzzer alone by some movement of both limbs,[2] *i.e.* gave a conditioned response. For example, subject He gave two conditioned responses on the 3rd day and eight responses on the 4th day. It will be seen that nine of the twelve infants in the experimental group gave conditioned responses to the buzzer alone on the third day. Only 11 of the 12 subjects were tested for spontaneous recovery. Of those who showed conditioning on the previous day, nine showed spontaneous recovery, and one infant, who gave no conditioned responses on the previous day, gave several such responses on the spontaneous-recovery day. In general, then, the results of the experimental group indicate that conditioning of the withdrawal response, its extinction, and the spontaneous recovery of the response on the following day, are obtainable in the under-ten-day-old infant.

The clearness of this picture is, however, clouded by the findings of the first control group, for the behavior of these subjects was practically identical with that of the experimental group, in spite of the fact that the subjects of Control Group 1 had never had a paired presentation of buzzer and shock. The results for Control Group 1 will be found in Table II, and may be read in the same manner as Table I.

TABLE II

SHOWING THE NUMBER OF RESPONSES TO THE BUZZER ALONE FOR CONTROL GROUP 1 ON THE THIRD AND FOURTH DAYS OF THE EXPERIMENT

Number of Responses	Subjects											
	Hi	Al	Ad	Pa	Ja	Th	En	Sm	Am	Kl	Fj	Wi
Day 3....	2	1	0	2	3	2	5	2	3	2	2	1
Day 4....	2	10	0	9	3	2	3	–	5	–	2	1

[2] Movement of both limbs was demanded as a criterion of the conditioned response since this bilateral movement was typical of the response to the unconditioned stimulus.

Ten out of the 12 infants responded to the buzzer, after the 36 shocks alone; all of these showed extinction, and of the ten babies who were tested on the fourth day, 8 showed spontaneous recovery. Needless to say, the slight differences between this control group and the experimental are not statistically significant. Conditioning, if you will, extinction, and spontaneous recovery are just as clearly demonstrated in this group as in the previous one.

It immediately occurred to the experimenters that these results might be due to faulty mechanics in the experimental set-up. It might be possible that, although the infant was enclosed in a supposedly sound proof cabinet, he was able to react to the sound of the vibrator on the inductorium itself, and this might become a conditioned stimulus which could be easily generalized to the buzzer. That this was not true was proved by including checks brought about by activating the inductorium alone without the application of shock. In no cases were any responses made to these checks.

There still remained the possibility that the responses to the buzzer alone given on the third and fourth days were due simply to maturational factors (*i.e.* increase in auditory acuity) rather than to any effects of the experimental routine. As will be recalled, Control Group 2 was designed to investigate this possibility. This group simply received two tests for responses to the buzzer, separated by two days. Of course, as in the other groups, no subject was included who responded to the buzzer on the first day. Only one out of the 12 subjects used gave any response to the buzzer on the second test. Applying the Chi Square formula we find that there is a statistically significant difference between this group and the experimental group and between this group and Control Group 1.

At this point it would perhaps be in order to consider the only other study we have found which investigated conditioning of the withdrawal movement in the under-ten-day-old infant, at the same time employing a control group similar to our Control Group 1. Wenger (**8**), using five infants in his experimental group, concluded that the movement of with-

drawal from shock in the infant could be conditioned to a tone. He ran a control of four subjects in which shock alone was given, followed by tests for responses to the tone (without any pairing of conditioned and unconditioned stimulus). This control group did give a number of responses to the buzzer alone. However, because the experimental group gave a few more such responses, Wenger concludes that there were essential differences between the two groups. An examination of his tables, however, indicates that this is not necessarily true, that there is no statistically significant difference between his experimental and control groups, and that his results are actually in accord with ours. Whatever differences there exist between the two groups may be attributed to sampling errors.

Discussion.—There seems, then, to be an essential similarity between our experimental group and our first control group, both of them showing evidences of 'conditioning,' extinction of the conditioned response, and spontaneous recovery, even though the situation for the control group was not a conventional conditioning situation. This similarity seems not to be due to artifacts of the experimental set-up, or to any maturational effects. There remains the problem of accounting for it. Two possible interpretations present themselves. These are highly speculative, neither proved nor disproved by the extant experimental evidence. However, they are perhaps worthy of consideration with a view to further investigation.

(1) Perhaps the continued stimulation which results in the leg movement characteristic of response to shock serves to sensitize and make this response prepotent. The muscles involved in the response are, then, in a state of readiness, and may be tripped off at almost any stimulation. Highly analagous to this in the field of infant behavior are the findings of Pratt, Nelson, and Sun (**6**), Jensen (**3**), and others, which indicate that nearly any stimulus will set off the sucking response, supposedly a prepotent response in the neonate. Possibly analagous in the field of adult conditioning is the phenomenon of sensitization (Bernstein, 1). If this interpretation of sensitization is accepted, it would seem that our

infants are not showing conditioning in the ordinary meaning of the term. However, before sensitization can become a complete explanation, we must know more concerning the nature of this phenomenon. The fact that spontaneous recovery existed—that these withdrawal responses to the buzzer appeared as late as twenty-four hours after the last shock, sometimes indeed with greater frequency than in the tests immediately following shock—certainly implies that sensitization is a relatively long-lasting thing.

(2) Perhaps conditioning in the ordinary sense is operating after all, but we, as experimenters, are not clearly recognizing the nature of the conditioned stimulus. Perhaps the infant, in responding to the conditioned stimulus, is not responding to the sound of the buzzer as such, but merely to the buzzer's characteristic of being a sudden change in the environment. If we analyze the nature of the shock, we may see how the infant has become so conditioned, *i.e.* conditioned to respond to a sudden change in the environment. The shock has two characteristics; it is not alone a pain-giving and response-producing stimulus, but it also possesses the character of being a sudden, sharp change in the environment. That is, first there is no shock; then, suddenly, there is shock. This sudden change in the environment is the conditioned stimulus, and it is always followed by pain, the unconditioned stimulus. Thus, the organism becomes conditioned to a change in the environment, and the buzzer, which is also a change in the environment, may now act as a conditioned stimulus. If this were true, we would expect many other stimuli which are effective to the infant also to set off the conditioned response, and this, of course, could easily be attacked experimentally. The interpretation implies that the infant does not easily make discriminations between specific stimuli, and reacts instead to major changes and characteristics—a view which is also supported by the findings of Pratt, Nelson, and Sun (**6**).

These theoretical interpretations may be of some value in suggesting problems for future attack. In any case, our data seem to have reopened the question of conditionability of the newborn infant. The results would indicate that, without

further experimentation, it cannot be stated that the withdrawal movement in the neonate can be conditioned, in the ordinary sense of that term. Modifiability, however, is clearly shown, and conditioning of some sort may be operating. The investigation points incidentally to the need for adequate controls in conditioning experiments, especially in those which use lower organisms as subjects.

Summary.—To investigate the problem of conditionability of the neonate, three groups, consisting of 12 under-ten-day-old infants, were studied, the conditioned stimulus being the sound of a buzzer, the unconditioned stimulus a shock to the sole of the foot.

Members of the experimental group received 12 paired stimulations a day for three consecutive days. They were then tested for conditioned responses to the buzzer alone, the responses were extinguished, and on the following day tests for spontaneous recovery were made. Members of Control Group 1 received 12 shocks alone per day for three consecutive days. Like members of the experimental group, they were tested for responses to the buzzer alone at the end of the third day and on the fourth day. Members of Control Group 2 were tested for responses to the buzzer alone on the first day; then, two days later, during which time no stimulations were given, they were again tested for responses to the buzzer alone.

Results showed that the experimental group gave evidence of conditioning, extinction, and spontaneous recovery. Control Group 1 gave almost identical results, although members of this group had never had a paired presentation of conditioned and unconditioned stimulus. Only one subject in Control Group 2 responded to the buzzer alone on the third day.

Conditionability (in the ordinary sense of the term) of the neonate, and incidentally also of the foetus, is thus not unequivocably supported. Two possible theoretical interpretations are advanced.

(Manuscript received October 24, 1939)

References

1. Bernstein, A. L., Temporal factors in the formation of conditioned eyelid reactions in human subjects, *J. gen. Psychol.*, 1934, **10**, 173–197.
2. Dockeray, F. C., and Valentine, W. L., A new isolation cabinet for infant research, *J. exper. Psychol.*, 1939, **24**, 211–214.
3. Jensen, K., Differential reactions to taste and temperature stimuli in newborn infants, *Genet. Psychol. Monog.*, 1932, **12**, Nos. 5–6.
4. Marquis, D. P., Can conditioned response be established in the newborn infant? *J. genet. Psychol.*, 1931, **39**, 479–492.
5. Pratt, K. C., The neonate, (In) Murchison, Carl (editor), *A handbook of child psychology*, Second ed. rev., Worcester, Mass.: Clark Univ. Press, 1930, pp. xiii + 956 (pp. 163–208).
6. Pratt, K. C., Nelson, A. K., and Sun, K. H., *The behavior of the newborn infant*, Columbus, Ohio: Ohio State University Press, 1930, pp. xiii + 237.
7. Wagner, I. F., The establishment of a criterion of depth of sleep in the newborn infant. *J. genet. Psychol.*, 1937, **51**, 17–59.
8. Wenger, M. A., *An investigation of conditioned responses in human infants.* University of Iowa Studies in Child Welfare. Studies in Infant Behavior III, **12**, No. 1, 9–90.

THE TIME ORDER ERROR IN SUCCESSIVE JUDGMENTS AND IN REFLEXES: II. AS A FUNCTION OF THE FIRST STIMULUS OF A PAIR

BY HELEN PEAK

Randolph-Macon Woman's College

In the first paper of this series (**6**), judgments of the loudness of the second tone of a pair and reflex reactions of the eyelid to the same tone were both shown to be functions of the intensity of the first stimulus and the interval between stimuli, and it was indicated that these reactions, J_2 and R_2,[1] bore similar relations to the variables mentioned. It was suggested on the basis of these results that the positive time error would be accounted for, in part, by the observed decrement in J_2 which increased with the intensity of S_1.

It is the purpose of the present paper to discuss the data upon the judgments of the *first* stimulus of the pair, for changes in J_1 as well as J_2 must play a part in producing the time error, though heretofore no such analysis has been made of judgments of both members of a pair. Since the judgment of S_1 occurs after both S_1 and S_2 have been presented, J_1, like J_2, may be affected by pairing, and the judgments, whether in absolute or relative terms, will in that case be a product of the mutual effects of the two stimulation processes as well as their interactions with the same or different background conditions in the organism making the judgments. This possibility suggests the importance of investigating the presence or absence of concomitant variation in J_1 and J_2, and the results of this part of the study are also to be found in this paper. A third paper considers these data and those of the first paper in relation to various theories of the time order error (**8**).

[1] As in the first paper, S_1 is used to refer to the first stimulus of the pair, S_2 to the second, R_1 to reflexes to S_1, R_2 to reflexes to S_2, J_1 to judgments of S_1, and J_2 to judgments of S_2.

Apparatus and Procedure

The first paper of this series has described the apparatus and procedure employed in securing the data discussed (**6**). The procedure involved judging, on a 9 step absolute scale, the loudness of both members of a pair of tones furnished by a 1000 cycle tuning fork oscillator and earphones. The first stimulus of a pair varied in intensity from trial to trial, four different intensities being used. The second stimulus remained constant throughout, and was equal in intensity to the largest intensity of the first stimulus. Single stimuli of the same intensity as the first of the paired stimuli were interspersed in the series. The onset of the first stimulus and of the second were separated by intervals of 177 or 285 ms. and each tone lasted 50 ms. Lid responses to these same tones were recorded photographically. Data from 5 subjects are reported.

Results

1. *The Judgments of S_1 Paired and Single*

In Table 1 are found the average judgment values for each intensity of S_1 at the two stimulus intervals, along with the

TABLE 1

Average Value Loudness Judgments of S_1

		J_1—177 ms.				J_1—285 ms.			
Subject	Intensity S_1 →	1	2	3	4	1	2	3	4
B	*M*	1.8	2.1	2.8	8.9	1.2	1.5	2.6	9.0
	N	47	47	48	44	47	47	49	47
	σ	0.6	0.7	1.0	0.3	0.5	0.5	0.7	0.3
W	*M*	1.1	1.2	3.2	8.0	1.1	1.3	2.5	8.5
	N	48	47	48	49	49	48	49	46
	σ	0.6	0.5	1.9	1.1	1.3	0.8	1.3	0.9
S	*M*	2.0	2.9	4.4	7.5	1.5	2.3	3.5	8.4
	N	73	72	75	75	72	73	74	73
	σ	1.5	1.3	1.7	1.2	0.7	1.0	1.1	0.6
E	*M*	2.7	3.9	5.4	7.1	2.3	2.4	4.4	8.1
	N	40	40	39	41	34	35	37	35
	σ	2.4	2.1	1.9	0.9	1.1	1.0	1.9	0.8
D_2	*M*	1.0	1.1	2.0	3.5	1.0	1.0	1.6	4.0
	N	1	18	30	31	5	25	31	32
	σ	0	0.2	0.8	1.8	0	0	0.6	0.8

number of cases, and the standard deviation of the distributions. Table 2 contains the individual averages of judgments of different intensities of S_1 when the stimulus was single, not paired. The judgments of S_1 alone made during 177 and 285 ms. series have been averaged together since they showed no significant differences.

The following points are to be noted in a comparison of these Tables. (1) At the first three intensities of S_1, the values are greater for J_1 paired than for the unpaired J. This is true in every case, except for W at 177 ms. ($S = 1$), and for the same subject at 285 ms. ($S = 3$). (2) At the largest intensity there is a tendency for the paired J_1's to be smaller than judgments of single stimuli. This tendency is not wholly consistent, however, and occurs only for subjects W at 177 ms., S at both intervals, E at 177 ms., and D at 177 ms. In all other comparisons the paired J_1 was either greater or equal. (3) The J_1 values are larger on the whole at the 177 ms. interval than at 285 ms. except again when S_1 has a value of 4. Subject W constitutes the only exception to this tendency.

The degree of consistency in the direction of differences shown by the averages of Table 2 is indicated in Table 3.

TABLE 2

LOUDNESS JUDGMENTS OF SINGLE STIMULI

(*177 and 285 ms.*)

	Intensity $S_1 \rightarrow$	1	2	3	4
B	*M*	1.0	1.1	2.1	8.8
	N	17	21	19	19
	σ	0	0.3	0.7	0.8
W	*M*	1.0	1.1	2.6	8.4
	N	20	22	19	21
	σ	0.1	0.3	0.9	0.5
S	*M*	1.0	1.0	2.0	8.5
	N	23	23	24	25
	σ	0	0.2	0.6	2.0
E	*M*	1.0	1.3	1.7	7.6
	N	16	15	14	13
	σ	0	0.3	0.4	1.3
D_2	*M*	1.0	0.7	1.1	4.0
	σ	0	0.5	0.4	0.7

Here, as in the comparisons of the first paper, the data have been fractionated. Averages of the judgments made within each half session have been calculated and a comparison drawn between the judgments of paired and unpaired stimuli

TABLE 3

COMPARISON SESSION AVERAGES J_1 (PAIRED) AND J (SINGLE)

		177 ms.				285 ms.			
	Intensity $S_1 \rightarrow$	1	2	3	4	1	2	3	4
B	$J_s < J_1$	100%	100%	60%	10%	20%	100%	70%	20%
	$J_s > J_1$	0	0	40	80	80	0	20	70
W	$J_s < J_1$	60	80	100	20	40	60	40	60
	$J_s > J_1$	0	0	0	20	60	20	0	20
S	$J_s < J_1$	100	100	100	0	75	100	86	43
	$J_s > J_1$	0	0	0	0	12.5	0	0	0
E	$J_s < J_1$	66	100	100	33	100	100	100	0
	$J_s > J_1$	33	0	0	0	0	0	0	0
D_2	$J_s < J_1$	*	0	100	0	*	0	100	100
	$J_s > J_1$		100	0	100		100	0	0

* See text (f. n. 3) for discussion of fact that when S_1 was value 1, D did not hear it as a separate tone. This also occurred part of the time at the next intensity.

within each of these sessions. Table 3 indicates the percent of these comparisons in which judgments of single stimuli were less than judgments of paired stimuli, as well as the percent of times when the converse was true. It will be seen that for all subjects the direction indicated by comparing the general averages of Tables 1 and 2 is supported by the sub-averages in a majority of the comparisons. These percents also indicate, as do the averages themselves, that the differences are greater at 177 ms. than at 285 ms. The number of comparisons involved is stated in Table 5 of the first paper (**6**).

2. *J_1 as a Function of the Intensity of S_1*

This relationship has not been plotted, since there are difficulties in the way of considering it as a true psychometric function. The difficulty arises, in the first place, from a lack of certainty that the intensity of the stimulus increases exactly in geometric steps. In the second place, there is a possibility of distortion in the distribution of the judgments produced

by the fact that the weakest and strongest tones used were part of the regular series. This may have affected the distances between judgments of stimuli of different intensities, particularly those at the extremes (1, p. 51). It is impossible, therefore, and unnecessary for the argument of the paper, to insist on conclusions about this point.

3. *Interrelations of J_1 and J_2*

Rank difference correlation coefficients between judgments of the first and second stimuli for each subject at each intensity and interval appear in Table 4. Inasmuch as the number of

TABLE 4

RHO'S (p) J_1–J_2

		177 ms.				285 ms.			
Subject	Intensity S_1→	1	2	3	4	1	2	3	4
B	P	.21	.15	.16	.60	.33	.12	−.15	.67
	PE_p	.10	.10	.10	.07	.09	.10	.09	.06
W	P	.31	.26	.33	.84	.33	.30	.07	.88
	PE_p	.09	.10	.09	.03	.09	.09	.10	.02
S	P	−.08	−.21	.19	.60	.02	−.21	.26	.54
	PE_p	.08	.08	.09	.05	.09	.08	.08	.06
E	P	.39	.33	.69	.85	.28	.47	.52	.79
	PE_p	.09	.11	.07	.03	.13	.11	.10	.05
D_2	P	*	*	.42	.80	*	*	.64	.54
	PE_p			.11	.05			.07	.09

* Too few J's for correlation—see f.n. Table 3.

values assumed by the correlated variables was often very small, the possibility arose that the correlations were spurious.[2] In order to check this, a scatter diagram was made of J_1 and J_2 values for each intensity, interval, and subject, and averages of the J_2 values falling in each J_1 category were calculated. The means were then examined in order to discover whether average J_2's paired with J_1's of each value tended to increase when the J_1's increased. A count was

[2] The limiting case of this effect is the absurd situation where one of the variables always has the same value, the median rank of the series. A p of 0.5 results. Tied ranks have the same effect in increasing amounts as their frequency increases.

made of the number of times an increase in average J_2's corresponded to an increase in J_1 in the scatter diagram. These results expressed in percent of total number of possible concomitant changes are found in Table 5, the results for all

TABLE 5

CONCOMITANT VARIATION—J_1 AND J_2

	177 ms.				285 ms.			
Intensity S_1 →	1	2	3	4	1	2	3	4
Percent increases in J_2 averages corresponding to increase in J_1 value..	33	47	66	93	44	43	65	100

subjects being included. When the values exceed 50 percent, this indicates that over half of the changes in J_1 values were accompanied by change in the same direction in J_2 averages. As the values approach 100 percent, the relationship between J_1 and J_2 becomes increasingly positive. It will be seen that the direction of differences from one intensity to the next is the same by both criteria and the following conclusions appear to be justifiable when both sets of data are considered. (1) There is an increase in the amount of positive relationship between J_1 and J_2 as the intensity of the first stimulus increases. (2) There is no correlation, or a slightly negative correlation at intensity 1 and 2. (3) There is marked positive correlation at the largest intensity of S_1.

4. R_1 *as a Function of* S_1

The most interesting point to be observed here is the fact that many of the stimuli used were below the reflex threshold, though they were in no case small enough to be imperceptible. Table 6 shows the percent of trials on which lid responses were not present following various stimulus intensities. Since there were no differences in results at 177 and 285 ms. intervals, they have been included in the same Table. Even at intensity 3, stimuli were occasionally sub-threshold. This brings out a further difference in the reflex and judgment functions, a difference in sensitivity levels. The stimuli used,

which were always above the threshold for perception, were very frequently sub-threshold for reflexes. This fact has been made use of in the theory of the first paper.

There are so few responses at intensity 2 that data of sufficient quantity for consideration are available only at intensity 3 and intensity 4. It is, therefore, impossible to compare the psychometric functions of J_1 with R_1, or R_1 with reflex functions previously published (**4**).

TABLE 6

PERCENT TRIALS WITH S_1 SUB-THRESHOLD FOR R_1

Subject ↓	Intensity			
	1	2	3	4
B	100	90	53	0
W	85	70	3	0
S	100	87.5	12.5	0
E	100	84	7.5	1.5
D_1	81	64	0	0
D_2	58	24.5	6.5	0

5. *Reflexes to S_1, Paired and Single*

It is possible to look for changes in the reflex as a result of pairing, like those found for the judgments. It is not surprising to find, however, that at the intervals here used, R_1 is unaffected by the stimulus which follows. The interval is evidently too long for summation effects and since the first reflex is over and done with before the second stimulus occurs, there can be no such effect as that noted in the judgments which occur after both stimuli are delivered. This appears to be one of the chief differences between the R and J functions.

6. *Interrelations of R_1 and R_2*

An examination of the interrelations between the first and second reflexes has been made, although the results are limited in value by the infrequency of the responses to the weaker intensities. It can be said, however, that at intensities 3 and

4 for all subjects at both intervals, there are small positive correlations between R_1 and R_2 and that, on the average, the correlation is greater at intensity 4 than at intensity 3. At 177 ms., following S_1 of intensity 3, the median *rho* is +.05, at intensity 4, +.22. At 285 ms., following intensity 3, the median *rho* is +.22, at intensity 4, +.27. In one subject, D, who showed enough responses in the lower intensities for correlation purposes, the increase in the value of *rho* is fairly consistent with the increase in intensity for both intervals, the values at 177 ms. being +.05, +.37, +.06, and +.35, and at 285 ms., +.05, +.17, +.24, and +.30. These slight relationships take on added significance from the fact that they support the findings of an earlier study in which correlation between R_1 and R_2 was seen to increase with increasing stimulus intensity (**5**). It is also to be noted that the tendency suggests a parallel with that found for J_1–J_2 correlations. Those correlations were largest at the largest intensity of S_1.

7. *Correlations Between R_2 and J_2*

It is possible to inquire further into the relationship between judgments and reflexes by correlating the values of J_2 and R_2 within each stimulus intensity and interval, and for each subject. This was done with the hope that it might reveal possible interaction between J and R levels or the influence upon these levels of some common determinant which was not under control in the experiment. The values of *rho* vary between −.17 and +.66 with a median of +.26 and no consistent tendency for correlation to be higher at the larger intensities. Only one-third of the 40 coefficients of correlation are equal to as much as 4 times their PE's.

Limitations of Present Experiment

Before discussing the results just reported, the following limitations of the present experiment will be set down so that certain points will not be overlooked in any future comparisons between these and other data. Certain major differences in this and other experimental attacks will also be pointed out.

1. Judgment of both the first and second stimuli has not been made in previous experiments. The question may be raised whether J_2 follows the same functions under these conditions as when J_1 is not judged on an absolute scale. It can only be said that one subject, D, did make judgments on this basis before she was required to judge both stimuli. This small amount of evidence suggests that J_2 follows the same function under both conditions.

2. Previous experiments have not made a comparison between the judgment of paired and unpaired stimuli occurring within the same series. Future experiments should seek to discover the effect on judgment of the single stimulus of inclusion within a series of paired stimuli. This is crucial to the determination of possible effects of expectancy established by pairing, upon J_1 and J_2.

3. The present experiment has used a much larger range of intensities than is commonly employed. The wide range was necessary in a preliminary experiment which was interested chiefly in a comparison of the J_2 and R_2 functions, because the differential threshold of the reflex is large and requires wide differences in the intensity of S_1 in order to demonstrate the effect of that intensity upon the decrement in R_2. It is obviously necessary to explore the intervening stimulus values under present conditions. Furthermore, the stimulus range should contain intensities beyond those used in the experimental series. Failure to do this probably distorts the end values of the experimental series.

4. The experiment should be extended to include situations in which S_1 and S_2 are equal at other stimulus intensities. Also, exploration should be made of situations in which S_1 is greater than S_2. This is necessary in order to test out the limits of the relation between intensities of S_1 and decrement in J_2, and it will be an important bit of evidence in connection with a theory of equilibration to be discussed, which hinges on relative rather than absolute values of the stimuli.

5. Other time intervals should be explored, again in order that the limits of the present generalizations may be set, and

that changes in the reflex and judgments at longer intervals may be compared.

6. Finally, it is important that the stimulus values be stated with reference to a threshold. This was impossible because of the limitations of the apparatus. The exact relations of the stimuli should also be known, though the error in this respect is certainly small in the present experiment.

Summary and Discussion

The present paper is a continuation of the report of an experiment in which subjects judged paired tone stimuli which elicited lid reflexes. The first stimulus of the pair varied in intensity; the second remained constant and followed the first at either 177 or 285 ms. Single stimuli were interspersed in the series. The first paper has reported certain facts about the decrease in size of judgment values of a stimulus and amplitude of a reflex to the same stimulus when occurring second in a pair. The present paper has undertaken a consideration of the judgments of, and reflexes to, the first stimulus, the interrelation between pairs of judgments, between pairs of reflexes, and between judgments and reflexes. A summary of the similarities and differences in judgment and reflex functions follows.

1. In the first paper judgments of the second stimulus of a pair and reflexes to the same stimulus were shown to be similarly related to changes in the intensity of a preceding first stimulus, except at the largest stimulus intensity where there was a reversal in the J_2 function.

2. J_2 and R_2 were also shown to be similarly affected by increase in the interval between S_1 and S_2 and, as a consequence, it has been suggested that these aspects of R_2 and J_2, as well as certain similar findings for spinal reflexes, may be subsumed under the same generalizations regarding the development of inhibitory processes.

3. In the present paper J_1 and R_1 are both seen to increase with increase in intensity of the stimulus, but the psychometric functions cannot be compared because of the reflex threshold which is so high that few reflexes were obtained at the two

lowest intensities of S_1. The higher reflex threshold constitutes an outstanding difference in the two reactions.

4. The study of judgments of the first stimulus of a pair has revealed the further fact that J_1 paired tends to be greater than judgments of single stimuli except when S_1 has value 4. No such effect of pairing is apparent in the R_1 results. This adds a further difference in the reactions at these two levels of behavior. Such a result is to be expected in view of the frequently observed difference in time factors at different neural levels. Lorento de Nó (**2**) has demonstrated the existence of self-reëxciting chains in the nervous system and has suggested their role in facilitation and inhibition. It seems reasonable to suppose that, as the complexity of neural organization increases, the number and length of such reverberating systems will increase, with the consequence that neural after-effects set up by a stimulus should persist longer at complex levels than in simpler systems. It is significant that Lorento de Nó (**3**) and others have commented on the longer time values at cortical levels.

The shorter latency of reflexes as compared to voluntary and higher level responses is doubtless an additional factor of importance in the difference under discussion. Both these temporal characteristics of cortical functions, their greater persistence and their longer latency, would help to make possible an overlapping of the after-effects of S_1 and those of S_2.[3] The third paper of the series will discuss other possible explanations of the effect of pairing on J_1 (**8**).[4]

5. It should be noticed that, whatever may be the source of the enhancement of J_1 paired, it will tend to increase the positive time order error as much as does the inhibition of J_2, and

[3] In a note to be published (**7**) it will appear that subject D often failed to distinguish the presence of two separate tones at the lowest and second intensity of S_1, although neither tone was sub-threshold. This certainly indicates either a very long latency of response to S_1 and/or persistence of excitation following S_1 so that it overlapped with S_2 processes. There was never such overlapping in the reflexes.

[4] It is interesting to speculate whether such differences in the persistence of after-effects of stimuli at different neural levels may be related in any way to the relative ineffectiveness of backward conditioning as compared to the common occurrence of backward associations in the learning of verbal material. Longer enduring processes at the level involved in verbal learning might make possible backward associations.

must, therefore, be considered in the discussion of time error theories in the paper following.

6. Positive correlations have been shown to exist between first and second responses in both judgment and reflex at the larger intensities of S_1. At the two smaller intensities the correlations tend to be zero or slightly negative. Many factors which might be common to the first and second responses could be mentioned as possible sources of these positive correlations, though the reason for the fact that the correlations are positive only at the large intensities is not obvious. It might be supposed, however, that since additional nerve fibers become active when stimulus intensity is increased, responses to stimuli of the same intensity are more likely to involve the same pathways than are responses to stimuli of different intensities. As a consequence, with identical stimuli, transient sources of variation which would affect a first response would be likely to influence in some degree a second reaction following closely after and depending on the condition of the same neural tissue. It may be, then, that the positive correlations of J_1 and J_2 and of R_1 and R_2 at the large intensities are not due to the absolute size of the stimuli but to the fact that at intensity 4, S_1 and S_2 were equal in intensity. On the other hand, absolute size may be the determining factor and it may be that sources of facilitation or inhibition external to the responding system are 'tapped' only when the stimulus reaches a certain intensity. Other factors that might produce a correlation between J_1 and J_2 are discussed in the third paper (**8**).

7. Finally, small positive correlations have been observed to exist between J_2 and R_2. When interpreting these results it must be remembered that these correlations have been calculated from J_2's and R_2's occurring under a constant experimental condition. The low values are, therefore, not inconsistent with the conclusion that the functions of R's and J's are similar in certain respects. If, for example, stimulus intensity is varied systematically, then the size of R's and J's tend to vary in a similar fashion. When, however, no such factor is undergoing systematic variation, as was the case for

values correlated, changes in R and J must be determined by chance influences which may or may not affect both reactions. The present results indicate that only a small amount of variation in R_2 and J_2 is determined by the same chance variables, a result which needs checking by more cases, for the numbers for each coefficient are small. If these low correlations should be confirmed the result would suggest that the two reaction systems operate with considerable independence of each other and that any similar relations which they show to independent variables must be due not to interaction between levels, but to similar and relatively independent processes which are set going by the determining factors in question. Such facts should have important bearing on theories of organization of the nervous system.

(Manuscript received June 15, 1939)

References

1. Long, L., A study of the effect of preceding stimuli upon the judgment of auditory intensity, *Arch. Psychol.*, 1937, No. 209, pp. 507.
2. Lorento de Nó, R., Facilitation of motoneurones, *Amer. J. Physiol.*, 1935, **13**, 505–523.
3. Lorento de Nó, R., Inhibition of motoneurones. Reprinted from *Problems of Neural Physiology and Behavior*, 1936.
4. Peak, H., Intensity of stimulation and amplitude and amplitude and latency of the lid reflex, *Amer. J. Psychol.*, 1932, **44**, 785–788.
5. Peak, H., Inhibition as a function of stimulus intensity, *Psychol. Monog.*, 1936, **47**.
6. Peak, H., Time order error in successive judgments and in reflexes. I. Inhibition of the judgment and the reflex. This *Journal*, 1939, **25**, 535–565.
7. Peak, H., Graded inhibition at sub-threshold intensities (In preparation).
8. Peak, H., The time order error in successive judgments and in reflexes. III. Time error theories. *Psychol. Rev.*, 1940, 147.

A PHOTOGRAPHIC METHOD FOR STUDYING DISCRIMINATION-LEARNING IN CHILDREN

BY THEODORE A. JACKSON

Columbia University

Experimentation in discrimination-learning had been tedious and time-consuming until Lashley (**2**) gave us the jump-technique. That method was designed primarily for rodents and probably could not be used with very many different types of animals. The method described herein is designed for use with child subjects, and it too is a time-saver. Instead of using punishment to hasten the learning, economy is effected by speeding up the rate of giving the trials, and also by making it possible to use many subjects simultaneously. Description of the method will be divided into two sections: one dealing with the preparation and presentation of the stimuli, and the other with instructing the subjects and recording their responses.

Preparation of the stimuli involves making the stimulus patches, then photographing them. Presentation involves projecting them on a screen. These will be discussed in the order named. Since it is usually preferred to use black stimuli [1] on a white background, the most expedient procedure is to cut out white stimulus patches (Bristol board) and paste them on a black background (dull finish black cardboard). By thus reversing the black-white relations it is not necessary to make a positive print; the negative may be projected, and black stimuli on a white field will be obtained. The stimuli may be made any size that is convenient to measure, then the backgrounds are made according to the desired proportions. The size of the image on the film may be adjusted by varying the distance between the camera and the stimuli. The present method is designed primarily for the use of 35 mm film strips,

[1] If colored film is used then the reversal between stimulus and field should not be made.

but two inch slides or standard slides might be used. The advantage of the film-strip is that the presentation-order of the stimuli may be pre-arranged. This reduces to a minimum the possibility of confusion, and it eliminates the possibility of secondary cues from a slide carriage. Secondary cues from the film itself are also eliminated, for the same frame is rarely if ever repeated. It is much easier to turn up a film strip to the next frame than it is to change slides, particularly if, in the latter, one is required to keep track of the order of presentation.

Since there will be many different arrangements of stimulus patches to be photographed, and these must be made from the same distance and at precisely the same angle, it will be more convenient to use a holder for both the camera and the stimuli. Vertical models resembling modern enlargers are most satisfactory for the stimuli may be arranged in the horizontal plane. If there is a tendency for the stimulus patches to curl and thus cast shadows, a plate glass may be placed over them. It is important that the proper type of 35 mm camera be selected. It must have a good lens that may be focused up to two feet or less; it should have a ground glass attachment for focusing and composition. The least expensive camera that will satisfy these requirements is the Sept. Although it is no longer manufactured, second hand ones can usually be obtained through large camera exchanges for about $20. Expensive candid cameras with the proper attachments are also adaptable. Either single frame or double frame cameras are all right. There is no particular advantage in the larger size. The present trend seems to be to use single frames for film-strips. Positive film, because of its high contrast, is probably the best type of film to use. It should be used with contrast developer.

It is important to select a good projector if the stimuli are to be accurately presented on the screen. It also should have a good lens [2] and it should have at least a 200 watt bulb as a light source. Weaker projectors could hardly be used dur-

[2] Frequently the same lens may be used in the camera and the projector, thus saving the expense of two good lenses.

ing the day even with the window shades drawn. It goes without saying that the darker the room the greater the contrast between the stimuli and the background. If filmstrips are used, it is essential that the projector have a device for automatically adjusting the frames as they are turned up; otherwise much time will be lost. The screen, if used in the front of a large school room, should be at least four feet square. The "half-tone" type of screen surface gives better contrast. It should be placed at exactly 90° with respect to the line of projection. The distance between the screen and the projector should be such that the projected stimulus fields will just about fill the screen. It is most important to see that all subjects can view the entire surface. It is impossible with the present method to arrange that each subject get exactly the same view of the stimuli. That no appreciable error will occur because of the differences in distance and angle-of-view is well supported by the many studies of constancy.

The responses of the subjects are recorded by a system which requires the child to put numbered tickets into either of two ballot boxes. Each subject is given as many tickets (hat check stubs) as there are trials to be given in an experimental session. These are numbered consecutively beginning with No. 1. They are placed on a hook in front of the child so that he can get them off easily as they are needed. Two small boxes (two inches square and three inches high) are also placed in front of the child, one on his right and the other on his left. These ballot boxes have in their tops small slits of appropriate size for inserting the tickets. At each stimulus presentation, each subject is to place a certain numbered ticket in one of the two boxes before him. If the positive stimulus appears on the right-hand side of the screen the correct response for the child is to drop the ticket in his right-hand box, and vice versa for the left-hand side. Since the presentation of the stimuli is coordinated with the use of the numbered tickets, it is possible, after the trials are completed, to decipher the correct and incorrect choices made by each child. In order that one subject does not see what choice his neighbor makes, three-sided square screens, made by clipping

together two manila folders, were placed on each child's desk. The open side of the screen was toward the child and the ticket holder and the boxes were within it. A child could not observe the selection made by another except by very obvious craning. If this occurred the experimenter in the back of the room could easily detect it.

The method may be introduced to the children by some variation of the following instruction. "This is a kind of guessing game. Two pictures will be shown up here (experimenter points to the blank squares projected on the screen) and you are to guess which one you think is the right one. If you think the picture on this side (E points to the left-hand square) is correct, then put your ticket in the box on that same side of your desk. If you are not sure, guess. You are to put a ticket in one of the boxes each time a new picture is shown. Now let's practice a little before we begin. If the correct picture is here, which box will you put the ticket in? Show me by raising your hand on that side. (If all appear to understand, proceed to the first trial.) Now everybody get ticket No. 1. Here is the first pair of pictures. Look at them carefully and decide which one you think is the correct one, then drop the ticket in the box on that same side. (When all have dropped their tickets, turn up the film to the next frame which will indicate whether they made a correct choice. The square which contained the positive stimulus will now have the words "This Was the One".) Let's see the hands of those that got the first one right. Now everybody must try to get the next one right. Get ticket No. 2." This informal procedure may be continued through the first five trials. By that time the routine will be sufficiently well understood that no further instruction will be needed except in special cases now and then.

In using this method it has been found that the motivation stimulated by the verbal instruction together with the competitive element that spontaneously appeared was sufficient to carry our subjects through one hundred trials without apparent diminution of interest. The first time the method was used, material rewards were promised to all subjects who

worked well, but these could be distributed only after the training was completed. Although they were delighted to get their rewards(marbles for boys and beads for girls), it was obvious that there was no significant functional relation between the rewards and the learning. It had been thought that this means would be necessary to maintain the interest level. Since this novel type of work seemed to be interesting enough in itself, no material rewards were used in subsequent work. Satisfactory results were obtained under this condition. If however this method should be used with individual subjects, it would undoubtedly be wise to use a material reward, for the factor of social stimulation would be absent.

In order that all subjects understood the procedure perfectly, twenty-five preliminary practice trials have been used. The first five have been described above, and the next five were similar but were given without additional instruction. In the next five information regarding the correctness of their choices was omitted. This was done to prepare them for the omission of the check-up in the critical trials which followed the regular training. Trials 16 to 20 were like 6 to 10, and 21 to 25 were like 11 to 15.

The presentation time of the stimuli has been controlled in the following manner. Alternate quadrants of a twelve inch disk of stiff black cardboard were cut out. The disk was rotated, by means of a clock mechanism, at the rate of five revolutions per minute. It was placed so that the edge turned in front of the lens of the projector, and thus alternately intercepted the beam of light. In this way the stimuli were automatically timed for three-second exposures; intermissions were also of three-second duration. This exposure time allows the slowest child ample time to make his choice and drop his ticket. The routine used is as follows: Subjects get tickets; stimulus is presented for three seconds; intermission for three seconds; information picture is shown indicating whether choice was correct or not (three seconds); intermission during which time they get the next ticket. Although this method gives uniform exposure of the stimuli and maintains a fixed rhythm, it seemed too slow and inflexible particularly after

the subjects had learned the discrimination. All subjects, after a few trials, make up their minds and drop their tickets before the exposure time is up. They have to wait the remainder of that period as well as through the intermission before they are able to check-up on their response. Under this condition of delayed reward, training is less efficient and interest is dissipated. The following modification of the procedure, though it involves less uniformity of exposure time, may be better in many cases. After a given exposure is begun the subjects are instructed to make their choice, and drop their tickets, and then raise their hands as quickly as they can. When all hands are raised the film is turned up to the information frame. After a few trials of training on a given set of stimuli all subjects will be able to finish within two seconds. The information frame might be shown immediately, but for only about a second. If an intermission of two seconds is allowed during which the subjects get their tickets for the next trial, the complete cycle from trial to trial would require but five seconds. This is less than half the time required in the first method.

The method of giving information to the subjects, described in the foregoing, might also be modified. It is suggested that the stimuli be presented again along with the information. This might be done by having an arrow point toward the correct stimulus rather than using the words "This Was the One." Since any modification of the pattern of the stimulus or of the stimulus field would disturb perception of the interrelations between the stimuli, the arrow must be placed outside either background. It might best be placed above. This suggested modification will undoubtedly make any task easier to learn, because the subjects may see the stimuli while they are checking-up. This can hardly be regarded as a disadvantage, so the alternative procedure will probably be preferred.

The method has been used successfully with second and third grade pupils. It could be used with any children that can recognize numbers, and possibly even younger if the

experimenters take care that all subjects use the right numbered ticket each trial.

As in most discrimination work, the method has been described in terms of two stimuli, but any number from one to nine may be used. An adaptation of this procedure has involved the use of nine stimuli. It is suitable not only for discrimination-learning but also for studies which are placed under such headings as 'multiple choice,' 'transposition,' and 'concept formation.'

It has been indicated earlier that many more trials may be made with this than with the usual methods. The writer using the traditional method (1) was able to give about 25 trials in 20 minutes to a single subject. With the present method it should be possible to give 100 trials to 15 children in about thirty minutes. The latter compares with the former, in terms of subject-trials per hour, per experimenter, as 20 to 1. It may occur to some that speeding up the trials will somehow dilute the quality of the results if not invalidate them. The opposite is probably the case. If a child subject is forced to wait between trials his interest is certain to decrease. Trials should be given as rapidly as is consistent with accuracy and with the comfort of the subjects. The child should be kept busy, and only brief (perhaps two seconds) intermissions should be allowed. Although this method has been designed especially for research with groups of children, its speed makes it of considerable value for work with individuals. It is probably four or five times as fast in individual work as the traditional method.

The advantage of the present recording system over ordinary paper-pencil methods is that the subjects are unable to look back over previous choices and thus run the risk of becoming confused. Also it is probably easier to drop tickets in boxes than it is to use a pencil to make a mark in a certain square on a mimeographed sheet. This advantage makes the method as a whole adaptable for a wider age range.

Summary

A method of studying discrimination-learning in children has been developed. The stimuli are photographed on 35 mm film-strips and projected on a screen before a group of 10 to 20 subjects. They indicate their choices between the right-hand and left-hand stimuli by dropping tickets (hat check stubs) into either of two boxes placed on their desks. Information as to which is the correct stimulus is also projected on the screen immediately after the subject's response. A record of each child's choice for each trial is determined by checking the numbered tickets dropped into either of the ballot boxes against the presentation order of the stimuli. As compared with traditional methods of studying discrimination, this method will enable one to gather about twenty times as much data per unit of experimentation time.

(Manuscript received June 27, 1939)

References

1. Jackson, T. A., and Dominguez, K., Studies in the transposition of learning by children: II. Relative vs absolute choice with multi-dimensional stimuli, *J. exper. Psychol.*, 1939, **24**, 630–639.
2. Lashley, K. S., The mechanism of vision. I. A method for rapid analysis of pattern-vision in the rat, *J. genet. Psychol.*, 1930, **37**, 353–360.

A STUDY TO DETERMINE THE RELATIVE EFFECTIVENESS (VISIBILITY) OF RED, ORANGE, YELLOW, GREEN AND BLUE, UNDER CERTAIN SPECIFIED CONDITIONS

BY FRANKLIN O. SMITH

Montana State University

The present study was motivated by the desire to determine the most effective color, or colors, for road signs considering dimness of illumination due to cloudy days, fog, and twilight. The experiment included the relative effectiveness (visibility) of red, orange, yellow, green and blue when seen in very dim light. For the purpose of the experiment, Hering colored papers 26 × 34 mm were pasted on strips of cardboard 1″ by 3″. The colors were presented in pairs by the method of paired comparisons. Twenty-two subjects, students with normal vision in general psychology, were given three trials of ten pairs each, a total of 30 comparisons for each subject.

The source of illumination was a 25 Watt incandescent lamp which was reduced by means of a filter to .01 of the original illumination, or about .25 foot candle.

A second object of the experiment was to determine the effect of contrast on the visibility of the colors included in the study. To measure the effect of contrast, narrow strips of black, white and orange were pasted across each of the five colors; one set of five colors with black strips, one set with white, and one set with orange. In another set the word 'Stop' was printed in black across each color.

The results are presented in Table I. Column 1 contains the pairs of colors compared; column 2, the number of times each color was judged more effective than its paired associate; column 3, the number of times each color when contrasted with black was judged superior to the other color of the pair; columns 4 and 5 give similar measurements for white and

orange respectively, and column 6, the same for the word 'Stop.'

The results are presented in terms of ratios based on the number of times each color of a pair was reported as being seen more clearly than the other color of its pair. For example, in a total of 66 comparisons of red and blue, red was reported as seen more clearly 60 times and blue 6 times. The ratio is, therefore, 60:6 in favor of red. The term "effective" as used in the present discussion is equivalent to visibility and the two terms will be used interchangeably for the sake of

TABLE I

SUMMARY OF RESULTS SHOWING RELATIVE EFFECTIVENESS OF THE FIVE COLORS UNDER THE VARIOUS CONDITIONS DESCRIBED ABOVE

1	2	3	4	5	6
Red : Blue	60 : 6	55 : 11	—	—	38 : 28
Orange : Red	64 : 2	65 : 1	50 : 16	—	63 : 3
Orange : Yellow	38 : 28	—	62 : 4	—	40 : 26
Orange : Green	55 : 11	58 : 8	41 : 25	—	37 : 29
Orange : Blue	66 : 0	65 : 1	44 : 22	—	54 : 12
Yellow : Red	56 : 10	66 : 0	57 : 9	63 : 3	62 : 4
Yellow : Orange	—	50 : 16	—	—	—
Yellow : Green	53 : 13	63 : 3	—	61 : 5	35 : 31
Yellow : Blue	62 : 4	66 : 0	—	—	45 : 21
Green : Blue	61 : 5	65 : 1	—	—	49 : 17
Green : Red	50 : 16	65 : 1	35 : 31	49 : 17	55 : 11
Green : Yellow	—	—	66 : 0	—	—
Blue : Red	—	—	46 : 20	49 : 17	—
Blue : Yellow	—	—	66 : 0	58 : 8	—
Blue : Green	—	—	44 : 22	38 : 28	—

variety. The terms superior and inferior are used to indicate greater or less effectiveness.

To summarize the results as presented in Table I, the first observation to be made is that orange is the most effective color ranking above all the other colors (col. 2). It is most effective when compared with blue, the ratio being 66 : 0, or 100 percent. Orange was judged more effective than red 64 out of a total of 66 trials. It was relatively least effective when compared with yellow. Thus orange is seen to be relatively more effective when compared with the extreme colors of the spectrum, red and blue, which are therefore farther from orange in brightness, than when compared

with those colors which are nearer in brightness to itself, as green and yellow. Yellow is seen to be more effective than red, green, and blue. It is relatively most effective when compared with blue, in which the ratio is 62 : 4 or 94 percent. Its relative effectiveness is least when compared with orange as shown by the ratio of 28 : 38, or 42 : 5 percent. The difference in effectiveness, when compared with blue and with orange, corresponds to the difference in brightness of blue and orange. The relative visibility of yellow when compared with red is approximately the same as when compared with green.

Green is more effective than red and blue. It is most effective when compared with blue with a ratio of 61 : 5, or 92 percent, and least effective when compared with orange with a ratio of 11 : 55, or 16 : 7 percent, and only slightly higher when compared with yellow in which the ratio is 13 : 53, or 20 percent.

Blue is the least visible of all the colors, its effectiveness being least when compared with orange (ratio 0 : 66) and greatest when compared with green (ratio 5 : 61).

When these colors are contrasted with black (col. 3), their visibility is only slightly different from what it is when compared without contrast effect as in column 2. The relative effectiveness of orange and yellow is reversed, yellow showing greater visibility than orange. This may be explained by the greater contrast effect of black on yellow than on orange.

Contrasting these colors with white, however, reveals some marked changes in their relative effectiveness (col. 4). Orange is again the most effective color, the ratio being greatest for yellow, and least for green and blue. Orange, however, is considerably less effective in comparison with all the other colors when contrasted with white than when contrasted with black or without contrast effect.

Under the same condition, yellow is inferior to all the other colors in visibility, its average effectiveness in comparison with red, orange, green and blue being in the ratio of only 3 : 63.

When red yellow, green, and blue are contrasted with orange (col. 5), blue proves to be the most effective color when

compared with yellow; the ratio being 58 : 8, or 88 percent. It is clear that the contrast effect of orange on blue is greater than on yellow or green; hence the greater relative effectiveness of blue.

Column 6 represents the effectiveness of a meaningful word, 'Stop,' printed in black, when seen against various backgrounds of color. It is most effective when orange and yellow are compared with red, 95 : 5 percent, and 94 percent respectively. This agrees substantially with the results as represented in columns 2 and 3, that is, when the colors are compared without contrast or when contrasted with black bars.

Green ranks next to orange as a background on which the word 'Stop' was read, when green was compared with red (ratio 55 : 11, or 83 percent). When compared with blue its effectiveness as a background was indicated by the ratio of 49 : 17. In other words 'Stop' is read less effectively against a background of green when green is compared with blue than when it is compared with red. In terms of the relative brightness of these colors when compared with green, this is what should be expected.

TABLE II

Table II contains the total number of times each color was seen more clearly than its paired color for each of the four conditions described in Table I.

1	2	3	4	5	Total	Percent
Red	88	57	76	37	258	21.5
Orange	223	204	197	—	624	69
Yellow	199	245	193	132	769	64.8
Green	135	141	148	82	506	42
Blue	15	13	178	145	351	29

The results show that orange was judged more effective than its paired color 245 times out of a total of 300 comparisons, or 70 percent of the times. Yellow was a close second with 199 out of 300 comparisons, or 67 percent of the times. Blue and red are least effective, being seen more clearly than their paired colors only 15 and 22 times, or 5 percent and 7 percent respectively.

When these colors are contrasted with black, yellow is judged the most effective 245 times, or 80 percent of the total number of comparisons. Orange follows with 204, or 68 percent. Blue is again the least effective color with a score of 13, or 4 percent.

When the colors are contrasted with white, green becomes the predominant color with a score of 148, or 61 percent, and blue slightly less effective with a score of 178, or 59 percent. Yellow falls to the bottom of the scale with a score of 14, or 4.7 percent.

When contrasted with orange, blue was judged the more effective over its paired color 145 times, or 48 percent, while the second highest in rank was that of yellow, 132 times, or 44 percent.

If we consider the totals of all comparisons under the four conditions, it is evident that orange predominates with a total score of 624 out of a total possible score of 900, or 69 percent. (There were 300 fewer comparisons for orange than for other colors since orange was one of the contrast colors.) Yellow ranked second in effectiveness with a score of 769 out of a possible score of 1200, or 64 percent, while green ranked third with a score of 506, or 42 percent. Red was the least effective with a score of 258, or 21.5 percent.

Conclusions

When the colors included in the foregoing experiments are seen in very dim light without contrast effect, orange is the most effective color. Its advantage is greatest when compared with blue, next with red, then green, and least yellow. Red is superior to blue, but inferior to the other four colors. Yellow is more effective than blue, red, and green in the order here indicated, while green is more effective than blue, but ranks below the other four. Blue shows the least visibility of all.

(Manuscript received June 28, 1939)

APPARATUS FOR PRODUCING INTERMITTENT AUDIBLE PULSES

BY BERTRAM WELLMAN AND LEONARD CARMICHAEL

In many psychological experiments, it is desired to produce definitely timed pulsations of tonal stimuli. Mechanical switches of various types devised for this purpose have all proved in certain respects unsatisfactory. The apparatus here described makes it possible to control the onset and the cessation of the stimulus and thus to avoid 'clicks' and other secondary stimuli sometimes produced by mechanical devices.

The present apparatus can also be used in timing any psychological performance; that is, it may be employed essentially as a metronome with a wide frequency range and with great adaptability.

The electrical apparatus here described produces pulses of audio-frequency electric current of any desired duration from less than one hundredth of a second up to a minute or more, separated by intervals of any desired length within the same range. No moving parts are used, its action being entirely electrical.

The left hand portion of Fig. 1 diagrams a thyratron inverter designed to produce a pulsating current of the desired duration and interval, and approximately rectangular wave form.

The pulses from T_2 are conducted to the grid circuit of the push-pull audio frequency amplifier shown at the right of the diagram, and make the grids of T_3 sufficiently negative to stop the flow of current in the plate circuit of this tube. Thus in the interval between the pulses of T_2, T_3 acts as an amplifier, while during the conducting period of T_2 its amplification is zero.

Any desired audio frequency may be introduced into the input circuit of the amplifier, and the output will consist of a series of pulses of that frequency. It is essential to use a push-

pull circuit in the amplifier in order to avoid the loud clicks which would otherwise be produced by the sudden interruption of the plate current.

If a wide range of duration and interval values is desired, it is best to use multi-point switches to select the necessary values of R_5, R_6, C_1, and C_2. Thus R_5 and R_6 may each consist of a 10 megohm variable resistor (volume control type) in series with an eleven point selector switch arranged with ten 5 megohm grid leak resistors to give from 0 to 50 megohms in 5 megohm steps. C_1 and C_2 may each consist of six condensers

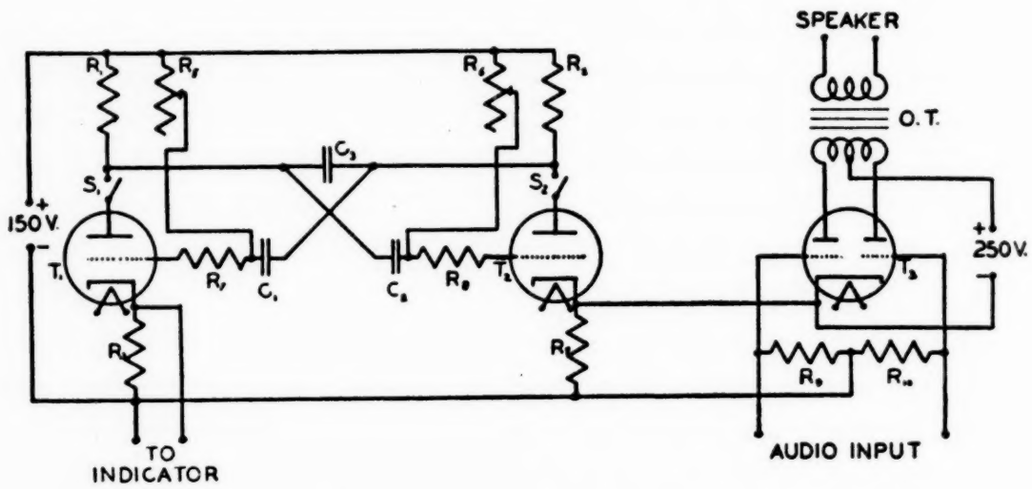


FIG. 1. T_1, T_2, type 885 thyratron; T_3, type 79 duo-triode; R_1, R_2, 10,000 ohms, 5 watts; R_3, R_4, 2,000 ohms; R_5, R_6, 0–50 megohms, variable; R_7, R_8, 150,000 ohms; R_9, R_{10}, 500 ohms; C_1, C_2, .001–2 microfarad variable; C_3, .001 microfarad; O. T., output transformer.

Note that the two power sources do not have a common negative, so either one may be grounded, but not both.

having values of .001, .005, .02, .1, .5, and 2.0 microfarads respectively, with a six point selector switch to permit the use of any one condenser. The duration in seconds of the pulse is approximately equal to the product of R_6 in megohms and C_2 in microfarads, while the interval between pulses is approximately equal to R_5C_1. Since the exact values of the duration and the interval between pulses are also dependent upon the characteristics of the thyratrons, the above formulas are only rough approximations, but they are useful in predetermining what values of R and C should be available if a more limited range is to be covered. Long times may not be secured by

indefinitely increasing R, since the grid of the tube itself conducts away the charge of the condenser, so if very long times are desired, larger condensers must be used.

In any work where oscillographic recording is used, it is convenient to record the pulses of T_1 as an indication of the 'on' period of the stimulus.

In starting the apparatus, be sure that switches S_1 and S_2 are open and remain open until the heater current of tubes has been on for at least 30 seconds, and the cathodes are at a red heat. Then close switches S_1 and S_2 successively. A single double-pole switch should not be used here, since if the switches are closed simultaneously, the inverter will not start. Control of the phase of the onset and cessation of stimulus may be obtained, if desired, by suitably impressing the audio frequency on the grids of the thyratrons.

(Manuscript received July 11, 1939)